Spring 2016

The Economic Rationale of Exhaustion: Distribution and Post-Sale Restraints

Ariel Katz, *University of Toronto, Faculty of Law*

2. The economic rationale for exhaustion: distribution and post-sale restraints

*Ariel Katz**

I. INTRODUCTION

The first sale doctrine, also known as the exhaustion rule of intellectual property (IP) rights, limits an IP owner's power to control the downstream distribution and use of its intellectual goods. Despite its common law origins, "an impeccable historic pedigree,"[1] and over a hundred years of adjudication, courts have never been able to draw the exact contours of the first sale doctrine or fully articulate its rationale.

Proponents of narrow (or no) exhaustion rules began harnessing insights from modern antitrust law and economics to suggest that, just as antitrust law has recognized the efficiency of some post-sale restraints and relaxed its hostility toward them, so should IP law permit their imposition and provide remedies for their breach. This chapter challenges this position. It shows that post-sale restraints can be beneficial, mostly in situations of imperfect vertical integration between coproducing or collaborating firms, when used to solve organizational problems that occur during the production and distribution phases or shortly thereafter. In such situations, the law should allow IP owners and those with whom they collaborate to contract around the first sale doctrine. Beyond such limited circumstances, however, the first sale doctrine promotes important social and economic goals: it promotes efficient use of goods embodying IP, guarantees their preservation, and facilitates user innovation, while minimizing transaction costs that otherwise might impede those goals. Therefore, rather than undermining it, the economics of post-sale restraints confirm the validity of the first sale doctrine and support its continued vitality.

This chapter focuses on the first sale doctrine in copyright law, but most of the analysis is applicable to other IP rights.[2] It also focuses mainly on U.S. law, yet most of its lessons can be applied elsewhere as well. The chapter proceeds as follows: section II briefly describes different existing and possible formulations of the first sale doctrine to set the stage for the discussion that follows. Sections III and IV present arguments in favor of the first sale doctrine and against it, while section V shows how those conflicting views can be reconciled and emphasizes how the first sale doctrine, properly understood, enables

* Associate Professor, Innovation Chair Electronic Commerce, Faculty of Law, University of Toronto. This chapter is based on an earlier article, Ariel Katz, *The First Sale Doctrine and the Economics of Post-Sale Restraints*, 2014 BYU L. Rev. 55 (2014). I thank Michael Stenbring for his research assistance.

1 Kirtsaeng v. John Wiley & Sons, Inc., 133 S. Ct. 1351, 1353 (2013).

2 Moreover, since the first sale doctrine is merely a manifestation of "a common-law doctrine with an impeccable historic pedigree," *Kirtsaeng*, *supra* note 1, 133 S. Ct. at 1363, the analysis can apply even more broadly to all restraints on alienation of property, whether or not IP rights are involved.

short-term post-sale restraints (where they are most likely to be beneficial) while guaranteeing the long-term benefits that exhaustion supports.

II. FIRST SALE DOCTRINE: DIFFERENT FORMULATIONS

Part of the difficulty in the debate about the first sale doctrine stems from ambiguity surrounding its meaning and scope: how strong it is, whether it is a mandatory or merely a default rule, and if it is a default rule, how sticky that default rule is. In order to begin answering these questions, it may be useful to recognize that there might be at least five possible formulations of the doctrine based on perceived strength.

Level 0: At one end of the spectrum lies the option of no first sale doctrine, meaning that it is up to the IP owner to decide whether the first-authorized sale or any subsequent sale would exhaust the IP right. This option is included not merely for the elegance of the model, but also because in some instances this is (or might be) the law.[3]

Level 1: A weak formulation views the first sale doctrine as a simple default rule: the IP owner's exclusive right is exhausted after the first *unconditional* sale. However, it may not be exhausted if the transaction is conditional and conditions were imposed by license, contract, or, possibly, even mere notice. Breach of any such conditions (by a contracting party, a licensee, or a purchaser who had notice of the restriction) would trigger liability for infringement.[4]

Level 2: Under a moderate first sale doctrine the first authorized sale still exhausts the IP right but the buyer may still be bound by contractual post-sale restraints.[5]

Level 3: A strong formulation of the rule treats the first sale doctrine as a sticky default rule.[6] This means that the first authorized sale exhausting the IP rights and attempting to work around exhaustion rules would be invalidated in the absence of a compelling case-specific explanation as to why the work-around should be upheld.[7]

Level 4: Under the strongest formulation of the first sale doctrine, not only does the sale exhaust the IP right, but also all attempts to work around the doctrine would be held invalid, and might even constitute IP misuse[8] or a *per se* antitrust violation.

---

[3] See e.g. Vernor v. Autodesk, Inc., 621 F.3d 1102 (9th Cir. 2010) (holding that the owner of the copyright in a computer program can escape its limitations by "licensing" copies instead of selling them); *Capitol Records, L.L.C. v. ReDigi Inc*., 934 F. Supp. 2d 640 (S.D.N.Y. 2013) (finding that the doctrine does not apply to digital copies resold apart from the medium in which they are embedded); *Kirtsaeng*, *supra* note 1, 133 S. Ct. at 1376 (Ginsburg J., dissenting) (regarding copies made outside of the U.S.).

[4] See e.g. Herbert Hovenkamp, *Post-Sale Restraints and Competitive Harm: The First Sale Doctrine in Perspective*, 66 N.Y.U. Ann. Surv. Am. L. 487, 541 (2010).

[5] See Brief of Amicus Curiae United States in Support of Petitioners, Quanta Computer, Inc. v. LG Elecs., 553 U.S. 617 (2008) (No. 06-937), 2007 WL 3353102, at *29.

[6] On sticky default rules, see generally Cass R. Sunstein, *Switching the Default Rule*, 77 N.Y.U. L. Rev. 106 (2002).

[7] This can be analogous to how the scope of fair use in copyright law develops, see Peter DiCola & Matthew Sag, *An Information-Gathering Approach to Copyright Policy*, 34 Cardozo L. Rev. 173 (2012).

[8] See Omega S.A. v. Costco Wholesale Corp., 776 F.3d 692 (9th Cir. 2015) (Wardlaw J., concur-

III. THE CASE FOR EXHAUSTION

The first sale doctrine has been justified on several grounds, economic and otherwise. Some arguments outline the ways in which the first sale doctrine benefits not only consumers, but also the public at large and even, in certain ways, creators and innovators. Exhaustion mitigates some of the deadweight loss that results from the grant of exclusive IP rights and the associated power to price intellectual goods above the competitive level, increasing access to and availability of intellectual goods mainly through secondary markets or parallel trade.[9] These markets make works sold through these channels more affordable, and therefore more accessible, and they also put competitive pressure on the prices that IP owners can set through their primary distribution channels.[10] Exhaustion also allows consumers who are unwilling or unable to pay the IP owner's market price to access the work through alternative channels such as rental, or borrowing from public institutions like libraries.[11] Perzanowski and Schultz offer another justification. According to them, exhaustion plays an important part in reducing consumer lock-in and thereby encourages platform competition; that is, competition between systems that can be used for different and changing applications.[12]

However, the effects of exhaustion on static efficiency are not unambiguously salutary. If IP owners respond to those competitive pressures by abandoning some markets or focusing only on "premium" high-margin markets, consumers who otherwise might be served will be worse off,[13] and because supply in the primary markets will be more limited, the quantities available through the secondary markets will be limited as well. That said, exhaustion may also benefit IP owners because it may assist them to implement *unofficial* price discrimination schemes, where official action may be more costly or even backfire on them. For example, students in a given geographical market may vary in their willingness and ability to pay for textbooks. Some can and would purchase a highly priced new textbook, while others cannot or would not. Ideally, the publisher would like to set different prices for each type, but it may be difficult to identify which student belongs to which type. Preventing the low-price types from reselling their books to the high-price types might prove difficult to enforce, even if such resale constituted

---

ring) (describing copyright misuse as an attempt to gain control over acts that fall outside the scope of the statutory grant).

[9] See e.g. Joseph P. Liu, *Owning Digital Copies: Copyright Law and the Incidents of Copy Ownership*, 42 Wm. & Mary L. Rev. 1245 (2001); R. Anthony Reese, *The First Sale Doctrine in the Era of Digital Networks*, 44 B.C. L. Rev. 577 (2003); Aaron Perzanowski & Jason Schultz, *Digital Exhaustion*, 58 UCLA L. Rev. 889 (2010) [hereinafter Perzanowski & Schultz, *Digital Exhaustion*]; Aaron Perzanowski & Jason Schultz, *Legislating Digital Exhaustion*, 29 Berkeley Tech. L.J. 1535 (2015).

[10] Ariel Katz, *The First Sale Doctrine and the Economics of Post-Sale Restraints*, 2014 B.Y.U. L. Rev. 55, 76 (2014).

[11] Reese, *supra* note 9, at 587–89.

[12] Perzanowski & Schultz, *Digital Exhaustion, supra* note 9, at 900 (exhaustion encourages platform competition by preventing lock-in, which occurs when "the costs of switching to a new vendor or technology platform are sufficient to discourage consumers from adopting an otherwise preferable competitive offering," by allowing consumers to recoup some of the costs they sunk into a platform through the resale of the platform itself, or the resale of applications which function on that platform, thereby making the switch to a new platform more affordable).

[13] Guy A. Rub, *Rebalancing Copyright Exhaustion*, 64 Emory L.J. 741, 763 (2015).

infringement. Moreover, high-price types might resent being charged a higher price for the same product, decreasing their own willingness to pay, or even increasing their willingness to consider obtaining pirated copies by reducing the moral cost of piracy.[14] Exhaustion offers the publisher a solution to this dilemma. First, it lets consumers sort themselves out. High-price types might be happy to pay for the ease, convenience, and value of a brand new book, which they might keep in their library after the course ends. Medium-priced types might value a brand new book, even one they do not intend to keep, knowing they can offset the high cost of buying by reselling it later. The low-price types might tolerate some inconvenience of obtaining a used book. While the publisher earns nothing from the sale of a used book, it earns it indirectly when it sells to the medium-price type at a price that otherwise they would not be willing or able to pay. When implementing official price discrimination schemes is costly, exhaustion can provide an effective second-best alternative, and because the secondary transactions occur irrespective of the IP owner's wishes, they are able to avoid some of the possible negative repercussions from officially condoning it.[15]

The effects of the first sale doctrine on dynamic efficiency are also ambiguous. To the extent that exhaustion constrains IP owners' market power and the profit that they may generate, it might reduce their incentive to invest in creating the intellectual good in the first place, making them, consumers, and society at large, worse off. However, the first sale doctrine can and does promote dynamic efficiency in certain ways. As discussed above, rather than impeding price discrimination, in some cases it might achieve similar, or even superior results, thus allowing the IP owner to earn higher profits and encouraging more investment in innovation rather than less.

Even if exhaustion increases competition from secondary markets, and even if this results in lower profit, it might actually accelerate the rate of innovation. Instead of investing in costly price discrimination schemes, some producers entice high-price types to continue paying premium prices by innovating and releasing new or upgraded products, which render last month's glitzy top-of-the-line product obsolete in comparison. The ability to resell the existing models makes even the medium-price types more inclined to pay a premium price and upgrade to the newest model.[16]

The first sale doctrine also contributes to dynamic efficiency by permitting secondary market channels to enable access to and use of intellectual goods and the ideas or technologies embedded therein, even if the IP holder ceases production or distribution of the good.[17] In a related vein, exhaustion serves an important purpose in preserving culture and knowledge in the long term by preventing the complete disappearance of works.

---

[14] Ariel Katz, *A Network Effects Perspective on Software Piracy*, 55 U. Toronto L.J. 155, 181 (2005).

[15] *Compare ibid*. at 179–86 (discussing how tolerating piracy allows software companies to implicitly price discriminate while avoiding some of the downsides of overt price discrimination).

[16] Margeurite Reardon, *Can Samsung Phone Trade-In Values Ever Match Apple's?*, CNET (February 7, 2014, 12:00 am PST), available at www.cnet.com/news/can-samsung-phone-trade-in-values-ever-match-apples/ (noting that "Savvy smartphone owners know that trading in their existing smartphone can help finance the purchase of their next device").

[17] Eric Matthew Hinkes, *Access Controls in the Digital Era and the Fair Use/First Sale Doctrines*, 23 Santa Clara Computer & High Tech. L.J. 685, 689 (2006).

While the first sale doctrine cannot be relied on for reprinting works that go out of print, it plays an important role in mitigating the potential cultural loss associated with such works. Exhaustion rules open up the possibility of a secondary market and assure that the artefacts embedding protected works remain available to the public over time.[18] Instead of discarding an item when it is no longer useful and becomes costly to keep or preserve, the first sale doctrine makes it legal to sell or donate a used copy of an intellectual good.[19] This way, the first sale doctrine enshrines preference for the garage sale over the garbage bin and for the library over the landfill.[20]

As Mulligan argues, the first sale doctrine reflects the *numerus clausus* principle, limiting the types of property rights exercised by producers and lowering the transaction costs of obtaining and using them.[21] Without the principle, significant costs would exist for consumers purchasing even simple low-cost products, such as books or CDs, because each item could be burdened with entirely different and fragmented terms and conditions.[22]

Exhaustion also safeguards consumer privacy.[23] Because they are able to use and transfer works without the permission of the producer, consumers are not subject to potentially intrusive monitoring mechanisms that can seem decidedly *1984*-esque.[24] Even without such aggressive monitoring, the simple fact is that "any system trying to answer the question of who can use what, when, where and under what circumstances will have to know where the object in question is, who is using it or whom it is being transferred to, and other details."[25] The first sale doctrine means that this kind of data collection is simply unnecessary.

Lower transaction costs and privacy also contribute to dynamic efficiency. Recent research on user innovation shows that important innovation often occurs outside the producer-firm. These studies challenge the producer-centric view of innovation, and the traditional belief that strong IP rights are necessary to spur innovation.[26] Strong IP

---

[18] Reese, *supra* note 9 at 592; see generally Diane L. Zimmerman, *Cultural Preservation: Fear of Drowning in a Licensing Swamp*, *in* Working Within the Boundaries of Intellectual Property (Rochelle C. Dreyfuss, Diane L. Zimmerman, & Harry First eds., 2010) [hereinafter Zimmerman, *Cultural Preservation*]; Hinkes, *supra* note 17, at 685; Margaret Jane Radin, *Regulation by Contract, Regulation by Machine*, 160 J. Inst. & Theoretical Econ. 1 (2004); Diane Leenheer Zimmerman, *Can Our Culture be Saved? The Future of Digital Archiving*, 91 Minn. L. Rev. 989 (2006) [hereinafter Zimmerman, *Digital Archiving*].

[19] Reese, *supra* note 9, at 607–8.

[20] Katz, *First Sale*, *supra* note 10.

[21] Christina M. Mulligan, *A Numerus Clausus Principle for Intellectual Property*, 80 Tenn. L. Rev. 235, 252 (2012–2013).

[22] Christina M. Mulligan, *Personal Property Servitudes on the Internet of Things*, Ga. L. Rev. (forthcoming 2016), Brooklyn Law School, Legal Studies Paper No. 400, at 32–36, available at http://ssrn.com/abstract=2465651; see also Molly Shaffer Van Houweling, *The New Servitudes*, 96 Geo. L.J. 885, 897 (2008); Perzanowski & Schultz, *Digital Exhaustion, supra* note 9.

[23] Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 Conn. L. Rev. 981 (1996); Julie E. Cohen, *Some Reflections on Copyright Management Systems and Laws Designed to Protect Them*, 12 Berkeley Tech. L.J. 161, 183–87 (1997).

[24] See Mulligan, *Personal Property*, *supra* note 22, at 46–47.

[25] *Ibid*. at 52.

[26] Carliss Baldwin & Eric von Hippel, *Modeling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation*, 22 Org. Sci. 1399 (2011).

rights may encourage innovation by producer-firms, but increase the cost of innovation that occurs outside those firms. Therefore, sensible innovation policies would also seek to maintain conditions that facilitate user-innovation, and should refrain from adopting rules that favor one model of innovation over others. The first sale doctrine supports user-innovation because it frees innovators who use existing intellectual goods from the need to obtain others' permission. Thus, exhaustion contributes to preserving the "Innovation Wetlands," a term coined by Andrew Torrance and Eric Von Hippel, to draw attention to the importance of various rules that foster "conditions that enable innovation by individuals to flourish."[27]

Finally, the first-sale doctrine also reflects constitutional norms of the rule of law and due process. Allowing IP owners to impose on others restraints beyond those that the statute specifically provides, to control what people can do with their own goods and then call upon the courts to enforce those restraints, delegates state power to private parties without any constraint and accountability. Such concerns influenced the U.S. Supreme Court's decision in some of the seminal exhaustion cases.[28]

IV. THE CASE AGAINST EXHAUSTION

The argument in favor of exhaustion set out in the previous section can be contrasted with the perspective of IP owners seeking greater control over their works, and, more recently, of some scholars, mainly from law and economics, seeking to highlight the efficiencies that can result from post-sale restraints. These theories draw in large part from developments in antitrust law and the economic theories that drove them. They contend that just as antitrust law has abandoned its earlier hostility to post-sale restraints, IP law should embrace them as well and allow IP owners greater, or even full, freedom to impose such restraints using the full arsenal of remedies available against infringers of IP rights, including third parties. In other words, the first sale doctrine should be narrowed, if not abolished altogether.

Post-sale restraints are a subset of a broader type of restraints, known as vertical restraints. They may restrict what a buyer can do with the goods she purchased—where she can resell them, to whom, at what prices, and whether she will have to provide pre- or post-sale services, repairs, warranties, etc. Antitrust law's attitude towards agreements imposing such restraints (e.g., between manufacturers and distributors) has seen remarkable changes throughout its history. Early antitrust law was as averse to vertical restraints as it was hostile to horizontal restraints,[29] and treated many agreements imposing verti-

---

[27] Andrew W. Torrance & Eric von Hippel, *The Right to Innovate*, 2015 Mich. St. L. Rev. 793, 798 (2015). See also Katz, *First Sale*, *supra* note 10; Pamela Samuelson, *Freedom to Tinker*, 17(2) Theor. Inq. L. (forthcoming 2016), UC Berkeley Public Law Research Paper No. 2605195, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2605195.

[28] Ariel Katz, *IP and the Rule of Law*, 17(2) Theor. Inq. L. (forthcoming 2016).

[29] Horizontal agreements are those entered between actual or potential competitors (i.e., firms at the same level of the distribution chain), whereas vertical agreements are those entered between firms at different levels of the distribution chain (e.g., manufacturer-wholesaler; wholesaler-retailer; retailer-customer).

cal restraints, such as exclusive dealing, tying, and resale price maintenance (RPM), as *per se* illegal.[30] Over time, however, antitrust scholarship began to recommend that horizontal and vertical restraints be treated differently because vertical restraints more likely enhance efficiency, prevent opportunism, and otherwise advance procompetitive outcomes than they are implemented for anticompetitive ends.[31] Jurisprudence followed scholarship, and since the late 1970s,[32] the Supreme Court has gradually abolished virtually all *per se* rules applying to vertical restraints.[33] The last bastion fell in 2007 in *Leegin*, which overruled an almost century-old *per se* prohibition on RPM.[34] This reflects the prevailing wisdom that frequently such restraints might be beneficial rather than harmful.

The argument relies in part on the important insight that unlike competitors who share a common interest in reducing competition among themselves at the expense of consumers, vertically situated parties do not share a common interest in reduced competition. Quite the contrary, a manufacturer benefits both when competition among its suppliers is intense and when market power among its distributors is limited.[35] Similarly, distributors would rather be free to deal with competing manufacturers upstream and competing retailers downstream than with monopolistic firms. In other words, the divergence of interests inherent in vertical relationships serves as a check on anticompetitive practices.[36]

A growing body of literature has identified a myriad of reasons supporting the proposition that they often serve procompetitive goals.[37] For example, vertical restraints may encourage dealers to invest in developing a local market (by advertising or other means) or to supply pre- or post-sale services (such as training or repairs). They may also be required to implement beneficial price discrimination schemes. As modern antitrust law has opened up to vertical restraints and their positive impact on competition, it has grown to tolerate agreements that govern and monitor them. In the case of goods in which no IP rights subsist, contracts and the mere threat of termination serve as the main tools for enforcing such restraints. But if IP rights can be relied on to enforce the restraints, the additional set of remedies could make those restraints more effective.[38]

---

30 Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 888 (2007).

31 *Ibid*. at 889.

32 Continental T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36 (1977).

33 *Leegin*, *supra* note 30, 551 U.S. at 900–904.

34 A nominally *per se* rule against tying still exists, but it has been watered down so significantly that it is questionable whether it is different from a rule of reason. See Ariel Katz, *Making Sense of Nonsense: Intellectual Property, Antitrust, and Market Power*, 49 Ariz. L. Rev. 837, 896 (2007).

35 Keith N. Hylton, Antitrust Law: Economic Theory and Common Law Evolution 253 (2003).

36 *Leegin*, *supra* note 30, 551 U.S. at 896 ("[I]n general, the interests of manufacturers and consumers are aligned with respect to retailer profit margins. The difference between the price a manufacturer charges retailers and the price retailers charge consumers represents part of the manufacturer's cost of distribution, which, like any other cost, the manufacturer usually desires to minimize.").

37 See e.g. Michael J. Trebilcock et al., The Law and Economics of Canadian Competition Policy 373–99 (2003).

38 Hovenkamp, *supra* note 4, at 492.

From this perspective, courts' stubborn adherence to the first sale doctrine undermines the efficacy of post-sale restraints, entrenches an inefficient IP rule, and preserves a relic from an era in which the economics of vertical restraints were not well understood.[39] Such courts fail to see that if IP rights subsist in the goods and can be relied on to enforce post-sale restraints, producers would have a more powerful and arguably more effective tool to enforce them. If post-sale restraints could be a good thing, the argument goes, then endowing IP owners with more powerful tools to impose and enforce them is a good thing too. Section IV.B below explains the basic flaw of this logical equation, but before doing that, let me present the argument in favor of enforceable post-sale restraints in greater detail in the context of parallel trade. This is a contentious phenomenon that attracts many of those arguments, though many of them are not limited to this context.

A. Parallel Trade

The term "parallel trade" (or "gray market") describes situations where goods sold abroad at a lower price are imported (or reimported) by an unauthorized dealer and compete domestically with the local authorized distribution system.[40] When such goods embed some protected IP, owners sometimes invoke IP law to ban the unauthorized importation. Whether IP law should be used for this purpose is subject to a heated debate and the actual rules are often inconsistent among nations, as well as between different IP laws within a nation (i.e., different rules for patents, copyrights, and trademarks) and even within a specific national IP law.[41] One question that plagues this debate is whether the IP right is exhausted only upon the first *domestic* sale authorized by the IP owner (national exhaustion) or whether the IP right is exhausted upon the first sale authorized by the IP owner regardless of the country in which it occurs (international exhaustion). Indeed, the difference between national exhaustion and international exhaustion was the dividing line between the majority and the dissenting opinions in *Kirtsaeng*.[42]

(1) Price discrimination

Proponents of national exhaustion often cite the benefits of international price discrimination when they argue that IP owners should be able to prohibit parallel trade. They maintain that because demand for an intellectual good (or the ability to pay) varies across countries, allowing producers to set different prices promotes both efficiency and social justice. Price discrimination may be more efficient because it may lead to higher output and greater incentive to innovate.[43] It may also be socially progressive because it might

---

[39] See e.g. Rub, *supra* note 13, at 754–59.

[40] See *Kirtsaeng*, *supra* note 1, 133 S. Ct at 1379.

[41] See e.g. Euro-Excellence Inc. v. Kraft Canada Inc., [2007] 3 S.C.R. 20 (Can.).

[42] *Kirtsaeng*, *supra* note 1, 133 S. Ct. at 1384 ("[I]n my view, [section 602(a)(1)] ties the United States to a national-exhaustion framework. The Court's decision, in contrast, places the United States solidly in the international-exhaustion camp.") (Ginsburg, J., dissenting).

[43] *Kirtsaeng*, *supra* note 1, 133 S. Ct. at 1390 (Ginsburg, J. dissenting); Katz, *supra* note 10, at 77; Rub, *supra* note 13.

help consumers with low willingness or ability to purchase goods, from which they would otherwise be excluded.[44] On this view, the first sale doctrine—a form of arbitrage that jeopardizes the producers' ability to price discriminate—leads to economically inferior and socially regressive results.[45]

Post-sale restraints may facilitate price discrimination in additional ways, for example, through the practice of tying, where a manufacturer sells Good A only on condition that the buyer use it with Component B, which can only be purchased from the manufacturer or from its designated suppliers. Tying can facilitate price discrimination when different buyers use different quantities of Component B with one unit of Good A. If Good A is worth more to the intensive users than to the less intensive users, tying Component B to Good A can effectively achieve the goal of discriminatory pricing for Good A. Component B, the tied product, serves as a counting device to measure how intensively Good A is being used.[46] For example, the demand for a patented machine may vary among different users. Instead of attempting to sell the machine to different users at different prices (and encountering the problems of determining in advance how much each buyer would be willing to pay and how to prevent those who paid a low price from reselling to those who paid a high price), the patentee may resort to tying: it can sell the machine at cost, or even for free "on condition that the unpatented staples used in the machine be bought from the patentee . . . Hence by charging a higher than competitive price for the staples, the patentee could receive the equivalent of a royalty from his patented machines."[47]

(2) Impact on local dealer investment

Antitrust scholarship has identified a variety of other benefits arising from post-sale restraints. Most of the benefits relate to guaranteeing local dealers a level of profitability in order to encourage them to make various investments in the distribution or the servicing of the product. Presumably, local dealers (distributers or retailers) would be reluctant to invest in promoting the product or in providing pre- and post-sale services if, after incurring these costs, consumers could buy the goods at a lower price from other dealers who free ride on their investments. To mitigate such concerns, a manufacturer may then impose various vertical restraints, such as territorial restraints (limiting the dealers to sell only within a designated territory) or price restraints (RPM). If the purpose of such restraints is indeed to guarantee investment and those investments are indeed necessary to increase output, then these restraints are efficient because they will result in more units sold and better services delivered.[48]

From this perspective, parallel trade is a cause for concern because it allows the parallel importer and the foreign dealer to free ride on the investments of the local dealer. Parallel trade, thus, could undermine the incentive to invest in building the local market and to

---

[44] See e.g. *Kirtsaeng*, *supra* note 1, 133 S. Ct. at 1390 (Ginsburg, J., dissenting).

[45] Rub, *supra* note 13.

[46] Ward S. Bowman Jr., *Tying Arrangements and the Leverage Problem*, 67 Yale L.J. 19, 23 (1957).

[47] *Ibid*. at 23–24.

[48] Edward Iacobucci, *The Case for Prohibiting Resale Price Maintenance*, 19 World Comp. L. & Econ. Rev. 71 (1995).

provide pre-sale and post-sale services, ultimately to the detriment of the local dealer, local consumers, and the manufacturer. To the extent that the first sale doctrine makes it more difficult to curb such parallel trade, it should be limited.

(3) Other benefits

Post-sale restraints may be used in distributing "positional goods," such as luxury brands, whose appeal lies in the exclusive status that their high prices confer.[49] Preserving the appeal of such goods requires an ability to maintain their high prices; parallel trade could undermine an effort by a producer of such a status good to maintain *different* high prices in different countries.[50]

Post-sale restraints may be used to achieve goals other than preventing arbitrage. For example, manufacturers sometimes argue that they need to exercise downstream control over the use and disposition of their products to ensure a guaranteed level of quality or safety.[51]

B. Limited Relevance of the Antitrust Insights to the Question of Exhaustion

From price discrimination to brand image, modern antitrust scholarship has identified a variety of benefits that post-sale restraints promote. These insights, which have contributed to the gradual erosion of antitrust law's hostility towards vertical restraints, have been carried over to the debates around exhaustion generally, and more specifically to debates around parallel imports. This section explains why, notwithstanding the importance of these insights, they have very limited bearing on the question of what model of exhaustion, national or international, IP law should adopt, and for deciding the scope of exhaustion more generally. In brief, those who rely on those economic insights to justify a legal position generally overshoot the mark.

Consider the following price discrimination argument. Price discrimination is good. Arbitrage frustrates price discrimination, therefore arbitrage is bad. An international exhaustion rule facilitates parallel trade, hence it supports arbitrage, which is bad. It follows that international exhaustion should be rejected and nations adopt only a national exhaustion rule. The flaw in the argument is threefold: first, it assumes that price discrimination is more efficient than uniform pricing, but no such general rule exists. Price discrimination may increase efficiency or it may not.[52] Second, it assumes that the benefits of legal rules facilitating price discrimination outweigh their costs; and third, it mismatches the symptom it identifies (arbitrage) and the remedy it prescribes (national exhaustion).

---

[49] Barak Y. Orbach, *The Image Theory: RPM and the Allure of High Prices*, 55 Antitrust Bull. 277, 279 (2010).

[50] See e.g. Nat'l Econ. Research Assocs. et al., The Economic Consequences of the Choice of a Regime of Exhaustion in the Area of Trademarks 104 (1998), available at http://ec.europa.eu/internal_market/indprop/docs/tm/report_en.pdf.

[51] Edward Iacobucci, *Tying as Quality Control: A Legal and Economic Analysis*, 32 J. Legal Stud. 435 (2003) (arguing that tying is not necessarily the optimal way for providing such assurances).

[52] Trebilcock et al., *supra* note 37, at 371.

If price discrimination is desirable, arbitrage is a problem, and IP law should be harnessed to prevent it, then the logical conclusion is to reject all exhaustion, not only international exhaustion. If the goal is preventing arbitrage, it seems odd to design rules that target only one type of arbitrage (cross-border), but remain agnostic towards all other forms. Efficient price discrimination requires an ability to sort consumers according to their willingness or ability to pay. Sometimes national borders would provide a close proxy to different consumer attributes, but in many cases the preferences and attributes of consumers within a nation will be just as varied as the preferences of consumers across nations. There is no general *a priori* reason to assume that international arbitrage is a problem that requires legal intervention, while domestic arbitrage is not.[53] If anything, transportation costs, regulatory differences, language, and various border measures limit international arbitrage irrespective of the IP regime, rendering IP law intervention less necessary. At the same time, the ubiquity of price discrimination schemes that occur domestically, notwithstanding the first sale doctrine, indicate that the supposed benefits of IP intervention are overblown. Thus, the price discrimination argument only explains why sometimes post-sale restraints may be justified, but it provides no coherent basis for choosing between national or international exhaustion.

Similar flaws plague most of the other legal arguments that rely on the benefits of vertical restraints to support limiting the first sale doctrine. Establishing and maintaining efficient distribution systems clearly benefits producers and consumers alike, and as I note in the previous sections, this goal often requires the imposition of enforceable post-sale restraints. If IP law should be asked to play a role in achieving better efficiency, national exhaustion provides only a very partial tool. For example, a manufacturer might assign exclusive territories or allocate types of customers to different dealers. Each of these dealers might need to make specific investments and might be reluctant to do so without being offered credible protection against free riding by other dealers or the manufacturer. Arguably, the first sale doctrine undermines the credibility of such guarantees because third parties might be able to obtain the goods and undercut the local dealer. If exhaustion rules should be curtailed to prevent that scenario, national exhaustion seems like an arbitrary and unprincipled choice because it targets only one type of arbitrage, not necessarily the most significant one. As a general matter, and in the absence of rules seeking to regulate particular industries under specific market circumstances to solve concrete problems,[54] the only principled choices are either no exhaustion at all or international (or indeed universal) exhaustion. Indeed, the logic of the economic arguments should result in empowering IP owners to impose all types of restraints.

As tempting as this position might be, the economic insights provide only limited normative guidance. They explain why there are circumstances under which post-sale restraints might be a good thing and, therefore, a sensible legal system might not seek

---

[53] Katz, *supra* note 10, at 86–88.

[54] It is always possible to conceive of specific situations where parallel trade exists to such an extent that the absence of legal tools to prevent it results in unambiguously undesirable results. See Katz, *supra* note 10, at 79–81.

to outlaw them altogether and adopt the strongest of the five formulations of the first sale doctrine mentioned above.[55] But other than that, the economic insights fall short of supporting the weakest formulations. More precisely they fail to: (a) establish that IP owners, as a general matter, should be entitled to rely on their IP rights to impose and enforce such restraints; and (b) identify the optimal legal tool for enforcing such restraints. Without more, the economic arguments do not tell us whether the power to control downstream use or dispossession of intellectual goods should be part of the IP bundle of rights or whether a more limited power should suffice.

Many of the economic arguments against exhaustion rely on a similar logic applied with respect to the "red wine" fallacy. Many people believe that drinking red wine confers protection against heart disease and hence is good for your health. This might not be true,[56] but even if drinking some red wine is good, it is false to conclude that drinking more red wine is better. Indeed, consuming a lot of red wine can be quite bad. Just as it does not follow from the general economic justification for IP rights that their scope and duration should be as broad and long as possible, it does not follow that granting greater powers to impose and enforce post-sale restraints might be a good thing. Like many questions of policy, economic analysis can be useful in assessing the implications of various alternatives and whether, relative to a certain baseline, the benefits that a proposed change might bring about will outweigh its costs, or vice versa. As the next section explains, the first sale doctrine (even in its strong formulations) does not prohibit all post-sale restraints, nor does it make it impossible to implement them effectively. Instead, it provides a general rule that balances the costs and the benefits that the power to control downstream uses of intellectual goods entails. Thus, the first sale doctrine assists IP owners in implementing post-sale restraints where their social benefits may likely be greater than their social costs. However, post-sale restraints limit their power where downstream control will likely yield smaller social benefits relative to the social harms that they might entail. Therefore, to choose among the formulations along the spectrum, from level 0 to level 4, it is not enough to point out that moving towards the weaker formulations (i.e., allowing IP owners to exercise greater control downstream) may yield some benefits, but it also requires accounting for the social costs that such a move would entail and whether the benefits outweigh the costs. In the following sections, I explain why this is unlikely.

V. RECONCILING CONFLICTING VIEWS

A. Exhaustion Does Not Mean the End of Efficient Distribution

Before we can choose the optimal strength level of the first sale doctrine, it may be useful to recall that the doctrine neither prohibits IP owners from achieving the benefits that post-sale restraints might bring about, nor does it totally impair their ability to do so. The first sale doctrine only ordains that as a matter of IP law, the owner's exclusive rights

---

[55] See II *supra*.

[56] Malcolm Law & Nicholas Wald, *Why Heart Disease Mortality is Low in France: The Time Lag Explanation*, 318 BMJ—Brit. Med. J. 1471 (1999).

do not include the power to impose the restraint, and therefore those rights cannot be relied on to enforce them. Even the strongest formulation does not prohibit all means of achieving the benefits of post-sale restraints.

Consider, for example, the case of a firm that owns a copyright in every country and is vertically integrated into production and distribution. Such a firm can achieve the same benefits from post-sale restraints by relying on its internal governance. It can implement internal rules and procedures to ensure that all its salespersons adhere to different retail prices that it chooses for different territories, conform to its quality standards, etc. The firm's ability to structure its preferred production and distribution system depends on how effectively its management can implement it. As a legal matter, corporate law, contract law, or labour law will be much more determinative of the firm's success, than copyright law's choice of any particular exhaustion rule.

While vertical integration may provide the firm considerable ability to structure its production and distribution system as it sees fit, it may have some shortcomings. First, the firm's internal controls do not bind anyone else, and therefore, once it sells the goods, they may circulate freely in a way that could undermine the goals that it sought to achieve. Therefore, the firm might wish to have some ability to control its goods even after it has sold them. Second, vertical integration may not always be possible or desirable. A firm might have an advantage in developing new intellectual goods, but not in producing, distributing, or providing post-sale service. Or it might be more efficient to collaborate with other players to provide some or all of these services than to do it alone. In such cases, the firm will not be able to rely on its internal command and control procedures but will need to resort to other legal tools, such as licensing agreements, to structure an efficient production or distribution system with restraints on some or all of its collaborators. If the law categorically prohibited any type of restraint, it would drive firms into greater vertical integration, even when it is inefficient to do so, or to forgo the benefits altogether. This may not be a good idea.[57]

Therefore, the firm might attempt to impose contractual restraints on its downstream collaborators or customers. But contractual obligations can only bind those who voluntarily assumed them, and therefore cannot bind third parties, such as the purchaser's own customers, users, or service personnel.[58] In theory, the manufacturer could require its contracting party to insert the same restrictions in the latter's contract with downstream buyers, but such obligations will be increasingly difficult to enforce. Moreover, remedies for breach of contract, which tend to be limited to expectation damages,[59] are usually weaker than remedies available for infringement of an IP right. Proponents of weak or no exhaustion rules usually invoke the shortcomings of contracts when they argue that IP owners should be able to use the more potent rules of IP law to impose and enforce the restraints.

But relying on contracts or any other legal tool also entails willingness to incur the costs of enforcing them and manufacturers can resort to various unilateral non-legal techniques

---

[57] Stephen M. Maurer & Suzanne Scotchmer, *Profit Neutrality in Licensing: The Boundary Between Antitrust Law and Patent Law*, 8 Am. L. & Econ. Rev. 476, 481 (2006).

[58] Hovenkamp, *supra* note 4, at 541.

[59] *Ibid*. at 539.

that make arbitrage more costly and less attractive. For example, they may refuse to sell to arbitrageurs or the buyers whom they suspect might immediately resell the goods. They might also limit the quantities sold to untrustworthy buyers or they might use different versions of the product to make the cheaper version less attractive or incompatible with the more expensive ones. Finally, they might delay the sale of the cheaper products and keep innovating and releasing newer products that render the older ones less attractive.

Moreover, resellers and arbitrageurs can only deal with goods that were made and put on the market by the IP owner or with its authorization, meaning that the supply of goods to the secondary market depends on supply to the primary one. Therefore, because the IP owner is the exclusive source of goods to the primary market, its decisions also influence the supply to the secondary market. Even though, at the margin, resold goods may compete with those sold by the IP owner, resellers operate (all things equal) at a cost disadvantage. Their supply of goods depends, albeit indirectly, on the IP owner whose decisions are made upstream. They may lack access to technical expertise, and they may not have access to the most lucrative distribution channels. Consequently, the IP owner has some control over its downstream competitors and some ability to prevent the growth of their market share at its expense. Granted, this power is indirect and imperfect, but in many cases, and especially when the lifecycle of goods is short, an ability to delay the growth of the secondary market, however temporary and imperfect, may be all that the IP owner needs to prevent exhaustion from having devastating consequences. In sum, exhaustion does not mean the end of efficient distribution systems, and in many cases IP owners can set them up even if they cannot rely on the remedial tentacles of IP law to enforce them.

B. Focusing the Debate on the Margin

Nevertheless, what may be good enough in many cases may not be good enough at the margin, and consequently, in theory, some efficient restraints will not be implemented unless IP owners can resort to the more potent tool of IP law to enforce them. Efficient business arrangements that depend on the enforceability of such restraints might be forgone. If so, then the inability to enforce the restraints would prevent efficient business schemes from being implemented.

Rational debate about the scope of exhaustion should focus on this margin. Proponents of weak exhaustion rules should be able to show that the marginal social benefits stemming from greater powers to control downstream uses outweigh the marginal losses, while supporters of strong exhaustion rules should be able to demonstrate that the additional marginal benefits are minor compared to the social losses that they might inflict. Unfortunately, we do not have the empirical data that would allow us to quantify and compare these effects. But fortunately, we can soundly predict their respective magnitudes by focusing more closely on the type of situations where the ability to impose restraints might be beneficial, the types of costs they inflict, and the likelihood that IP owners and others who might bear the costs of the downstream restraints will be able to agree on socially efficient outcomes. Such a closer look reveals that granting IP owners greater powers to impose and enforce post-sale restraints is likely to cause more harm than good.

All the economic justifications for enforceable post-sale restraints share two common features: (1) they primarily involve settings of joint-production and incomplete vertical integration; and (2) they seek to address short-term concerns. In other words, the economic

justifications for post-sale restraints describe them as solutions to concerns arising during production, initial distribution, or shortly thereafter, and as solutions to organizational problems between non-integrated or imperfectly integrated firms along the production and distribution chain. Whether post-sale restraints are imposed in the provision of pre-sale or post-sale services or to facilitate price discrimination, to maintain the status of a luxury good, or to encourage investment in building a distribution system, a closer look at the antitrust scholarship reveals that it has focused on the relationships between collaborating firms attempting to organize an efficient production and distribution system.

For obvious reasons, IP owners would prefer the ability to enforce post-sale restrictions through property mechanisms,[60] but while the availability of remedies for IP infringement reduces the costs of enforcing post-sale restrictions, such remedies may increase the associated social costs.

Modern antitrust teaches us that post-sale restraints are not necessarily harmful. They may actually be quite beneficial and necessary to organize sophisticated distribution systems when a manufacturer is not fully integrated into distribution and retail.[61] The following line of reasoning may thus be adopted: if post-sale restrictions are efficient, they should be enforceable. If enforcing them on the grounds of IP infringement is easier than on the grounds of breach of contract, an IP remedy should be available. Unfortunately, this line of reasoning is flawed. Remedies for infringement of an IP right may be more effective than those available for breach of contract. However, whether greater efficacy is desirable depends not only on the benefits of more compliance, but also on the costs that may be externalized to third parties.[62] Therefore, before concluding that greater enforceability is better, it is important to carefully understand why and against whom post-sale restrictions may need to be enforced.

C. Short-term and Long-term Costs and Benefits

The organizational problems that post-sale restraints might solve occur mainly at the early stages of the product lifecycle:[63] production problems disappear immediately after the good is produced[64] and distribution problems largely disappear upon the distribution of the good, or shortly thereafter. Notably, these problems cease to exist long before the IP rights expire (particularly in the case of copyright). Moreover, as co-producers, the firms are in privity, which enables them to rely on contracts for addressing many of the organizational problems associated with efficient distribution.[65] Further, because the standard remedy for breach of contract is only damages, and the plaintiff needs to prove actual

---

[60] See Part VI of Katz, *supra* note 10, at 89–94.

[61] See e.g. Hovenkamp, *supra* note 4, at 489.

[62] Henry E. Smith, *Toward an Economic Theory of Property in Information*, *in* Research Handbook on the Economics of Property Law 104 (Kenneth Ayotte & Henry E. Smith eds., 2011), available at http://ssrn.com/abstract=1712089.

[63] Ariel Katz, *Substitution and Schumpeterian Effects Over the Life Cycle of Copyrighted Works*, 49 Jurimetrics J. 113 (2009).

[64] *Ibid*.

[65] As opposed to organizational problems during the development and production stages, which contract law alone may not solve and which IP rights may help to ameliorate. *Ibid*. at 141–42.

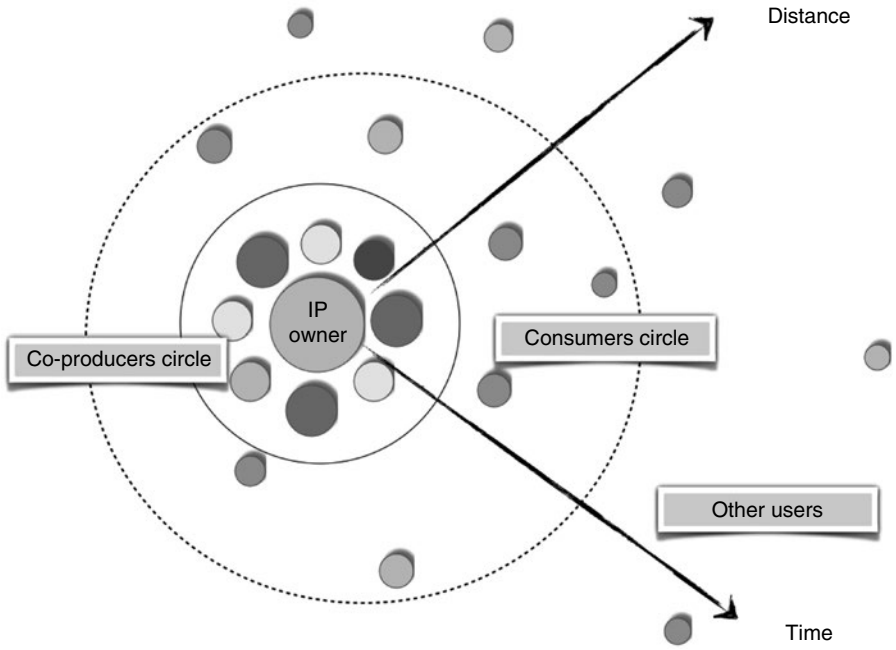

Figure 2.1 An IP owner, co-producers, and users

damage, a party bound by a contractually valid post-sale restraint may be able to put the good to better use when no damage can be shown or otherwise when the breach is efficient. Thus, while IP remedies may increase the enforceability of post-sale restraints, granting those additional powers yields marginally decreasing benefits.

Figure 2.1 represents the circles of relationship between an IP owner and interested parties through the lifecycle of an innovative good. The inner circle comprises the IP owner and the several firms that participate in the production or initial distribution of an innovative good. The second circle depicts the consumers who buy the goods produced in the inner circle. The third circle shows users who may at one point be interested in obtaining and using the innovative good but do not obtain it directly from the inner circle. Two vectors—distance and time—are also presented. Distance reflects the transactional proximity between the IP owner and the user. As the distance between the user and the IP owner increases, concerns about opportunism diminish. The time vector reflects temporal proximity between production, distribution, and use.

As noted above, the benefits of post-sale restraints are concentrated primarily within the inner circle. Hence, the marginal benefit from having enforceable restraints diminishes as we move along the vectors of distance and time.

In contrast, many of the benefits of exhaustion are long-term and distributed. Exhaustion frees users from the need to determine whether a product can be used without restriction and from the need to locate the IP owner and negotiate a license for reserved uses. The associated costs likely increase as we move along the distance and time vectors. When users contemplate combining various intellectual goods, those costs compound.

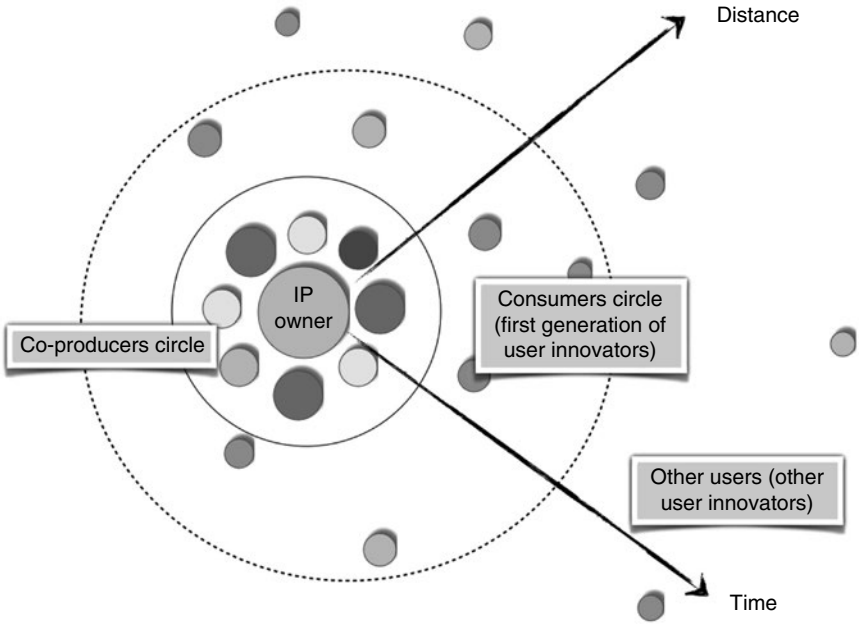

Figure 2.2 An IP owner, co-producers, and user innovators

Since users are not only consumers but also actual or potential innovators, granting IP owners an extended power (over time and distance) to restrain the use of goods embodying their innovation will impede users' ability to innovate or transfer the goods to others who might innovate. Figure 2.2 graphically depicts this assertion.

Moreover, the proximity between the IP owner and the players in the inner circles and the fact that they have already established business relationships for fairly well-defined goals imply that the transaction costs needed in order to structure efficient deals among themselves are relatively low. In contrast, the temporal and spatial distance between the IP owner and generations of consumers and future innovators, and the fact that the uses in which they might wish to engage are varied and sometimes unpredictable, imply that the costs of obtaining permission, if necessary, are likely to be fairly high.

It follows that it would make sense to allow IP owners to impose restraints where such restraints may result in significant marginal benefits, and limit such powers when the restraints will likely result in decreasing social benefits and increasing social costs. The first sale doctrine does exactly that. The law grants strong property rights that enable IP owners to enter into license agreements that can promote efficient production and distribution at the early and critical stages of production and distribution and to rely on the full range of the more potent remedies that IP law provides to enforce them. But the law, through the first sale doctrine, also limits IP owners' ability to exercise downstream control when the marginal utility of such control is diminishing. This way, the first sale provides room for efficient organization, but it also guarantees that trade will not be encumbered through durable, but often unnecessary, restraints later on.

VI. WHY EXHAUSTION SHOULD BE A STICKY DEFAULT RULE

The preceding discussion supports the existence of a first sale doctrine, but it does not yet fully answer how strong it should be. It explains why the grant of property rights that empower IP owners to impose restraints on third parties beyond the inner circle of co-producers (and after production or initial distribution) might cause more harm than good. It therefore suggests that the power to enforce post-sale restraints should not normally be part of the property bundle. Even under an assumption that post-sale restraints are efficient, they should generally be imposed and enforced as a matter of contract law, not property.[66] Assuming therefore that exhaustion is the default rule, the next question is how strong or sticky this default rule should be. Under what conditions should courts enforce deviations from the default and what are the grounds for invalidating such deviations? The issue is less about a categorical choice between IP remedies and contract remedies (and whether clever drafting can guarantee that a restriction is found to be a license condition rather than contractual covenant),[67] than about the design of default exhaustion rules.

I need first to reject the view that exhaustion provides no more than a simple default rule that can be easily modified by contract. Current case law tends to reflect this paradigm[68] and as a result, many cases are decided on the basis of relatively marginal legal questions, such as what constitutes a valid contract,[69] whether a first conditional sale pre-empts exhaustion,[70] whether there was a sale or just a license,[71] and whether restraints imposed by notice are sufficient.[72] Such decisions do not get to the heart of the problem.

The economic arguments in favor of easily contracting around exhaustion rely on the logic of the Coase Theorem that asserts that in the absence of transaction costs, bargaining will lead to an efficient outcome regardless of the initial allocation of property rights.[73] In a Coasian world, it does not matter whether the first sale doctrine exists or not because transacting parties will always be able to efficiently bargain about the rights to resale or otherwise use an item. If resale is efficient, the owner and the user will enter into a contract permitting it, and if it is not, the contract will restrict it, regardless of which party has the initial right to resale. If some consumers value the ability to resale the good more than others, then sellers would be happy to sell the goods with or without such rights at different prices.

While attractive, this policy prescription is flawed—the real world is not Coasian and

---

66 Katz, *supra* note 10, at 100. There might be some exceptions that could justify resort to IP remedies. *Ibid*. at 100–101.

67 MDY Indus. v. Blizzard Entm't., Inc., 629 F.3d 928 (9th Cir. 2010).

68 For more detailed examination of how the courts have dealt with these issues, including an in depth discussion of *ProCD*, see Part VII of Katz, *supra* note 10, at 100–109.

69 See e.g. ProCD v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996) (discussing the validity of shrink-wrap licenses and holding that they are valid, as long as the buyer can return the product after having an opportunity to read the terms).

70 Quanta Computer, Inc. v. LG Elecs., Inc., 553 U.S. 617 (2008) (leaving the question open).

71 Vernor v. Autodesk, Inc., 621 F.3d 1102 (2010).

72 UMG Recordings, Inc. v. Augusto, 628 F.3d 1175, 1180 (9th Cir. 2011) (distinguishing *Vernor* on, among other things, lack of acceptance of the restrictions).

73 Ronald Coase, *The Problem of Social Cost*, 3 J.L. & Econ. 1 (1960).

the world of IP is even less so.[74] The need for IP rights, and the need to define their limits, arises precisely because transacting around information, the subject matter of IP, is replete with all sorts of market failures.[75]

Therefore, we cannot assume that IP law provides only the baseline from which bargaining will necessarily, or even presumptively, increase social welfare.[76] This does not mean that contracting out of limitations on IP rights cannot increase social welfare; we have no reason to assume that the initial allocation is always optimal and we have grounds to believe that sometimes it may not be. However, it does not follow that whenever IP owners and the parties with whom they transact agree to work around the initial allocation of right, social welfare presumptively increases. The reason is simple: IP owners and their immediate transacting parties may rationally impose and agree to restraints that maximize their short-term private benefits, while ignoring the short- and long-term externalities that such restraints may generate. Since the terms of the transaction between the IP owner and her transacting party are unlikely to reflect what is socially optimal, contracting around the first sale doctrine should be met with a healthy dose of legal skepticism. Contrary to some views, the problem extends beyond mere notice.[77] Full notice of the restraint does not remedy the problem; it may only improve the bargaining between the parties. Notice does not address the externalities.

It follows that exhaustion should be treated as a sticky default rule. The law should not categorically invalidate any attempt to contract around the first sale doctrine, but it should also require those who seek to enforce the restraints to justify their efficiency and reasonableness before a court will uphold them. When the restraints purport to bind third parties or have long-term effects, the dose of suspicion should increase.

This view would be consistent with the common law historical treatment of agreements in restraint of trade. As Judge Taft explained in *Addyston Steel*,[78] the early common law treated all contracts in restraint of trade as wholly illegal but gradually, the common law began recognizing several categories of contracts in partial restraint of trade which, when reasonable, were generally upheld as valid.[79] A partial restraint of trade would be reasonable not only if it protects the interests of the parties, but also if it does not interfere with the interests of the public.[80] A reasonable restraint would be one that is merely ancillary to an otherwise legitimate contract and where it is:

---

[74] The term "Coasian" is misleading because it actually ignored Coase's main contribution, namely, that transaction costs are pervasive and important and that the law does matter. As Coase himself wrote, "[t]he world of zero transaction costs has often been described as a Coasian world. Nothing could be further from the truth. It is the world of modern economic theory, one which I was hoping to persuade economists to leave." See Ronald H. Coase, The Firm, the Market, and the Law 174 (1988).

[75] Katz, *supra* note 10, at 104–8.

[76] Except, perhaps, in the case of questions of who is the first owner, as opposed to the question of what this ownership entails. IP law clearly contemplates assignments of ownership, or the grant of licenses, which, by definition, presuppose the possibility that the owner is not necessarily the best exploiter.

[77] Hovenkamp, *supra* note 4, at 516–21; Rub, *supra* note 13, at 793–94.

[78] United States v. Addyston Pipe & Steel Co., 85 F. 271 (6th Cir. Court of Appeals 1898).

[79] *Ibid*. at 281.

[80] *Ibid*. at 282 (citing Horner v. Graves, 131 E.R. 284, 287 (1831)).

> inserted only to protect one of the parties from the injury which, in the execution of the contract or enjoyment of its fruits, he may suffer from the unrestrained competition of the other. The main purpose of the contract suggests the measure of protection needed, and furnishes a sufficiently uniform standard by which the validity of such restraints may be judicially determined. In such a case, if the restraint exceeds the necessity presented by the main purpose of the contract, it is void."[81]

Taft's judgment and the notion of "ancillary restraints" became the cornerstone for antitrust law's relaxed stance towards vertical restraints,[82] but its common law basis and logic are equally applicable to the issue at hand. Contracts limiting exhaustion will be reasonable if they are (a) ancillary to an otherwise legitimate business enterprise, and (b) do not impose restraints exceeding what is necessary to protect the interests of the one or both contracting parties and do not impose externalities that interfere with the interests of the public. The first sale doctrine, when coupled with rule that allows parties to contract around it if they can demonstrate that the restraint thereby imposed meets the same reasonableness standard, therefore rests on sound economic and legal grounds.

This brief discussion of the history of the common law doctrine of restraint of trade highlights a final crucial point for the formulation of optimal exhaustion rules: the question of burdens of proof. Under modern antitrust law, vertical restraints are subject to rule of reason analysis, meaning that the burden is on the plaintiff to show that an agreement in restraint of trade produces the requisite anticompetitive effect. Proponents of weak exhaustion rules, who uncritically apply the lessons from the development of antitrust law, maintain that IP owners too should be able to impose post-sale restraints unless it can be shown that those restraints result in a concrete harm to competition or to innovation.[83]

However, the logic of antitrust with respect to burdens of proof does not carry over to the context of IP. First, the goals and focus of IP laws are not the same as those of antitrust. As noted above, IP laws' own logic requires some stickiness to its allocation of rights between owners, users, and the general public. Second, in addition to invalidating the agreement, a finding of an antitrust violation may result in civil and potentially criminal liability. Generally, the rule of law mandates that a person cannot be held liable unless that person's liability has been established under the appropriate standard of proof. In contrast, strong exhaustion rules, as this chapter suggests, do not result in liability, but only in incapacity. An IP owner who attempts to impose an unreasonable restraint is not, as a matter of IP law, guilty of any offense or liable to pay damages, but is merely incapable of harnessing the court to impose its wishes on others. Because exhaustion, under this formulation, merely sets a limit on the scope of the IP owner's statutory grant in order to preserve larger public interests, rule of law considerations do not require any particular burden of proof.

---

[81] *Ibid*. at 282.

[82] Robert H. Bork, *Ancillary Restraints and the Sherman Act*, 15 Sec. Antitrust L. 211 (1959).

[83] Christina Bohannan & Herbert Hovenkamp, Creation without Restraint: Promoting Liberty and Rivalry in Innovation 393 (2011).

VII. CONCLUSION

Despite over a hundred years of adjudication, courts have never been able to draw the exact contours of the first sale doctrine or fully articulate its rationale. Recently, insights borrowed from modern antitrust law and economics have been invoked to provide a seemingly robust theoretical foundation for undermining exhaustion rules or narrowing their scope, thereby strengthening IP owners' control over downstream distribution and use of the goods they produce. It has been suggested that just as antitrust law has recognized the efficiency of post-sale restraints and relaxed its hostility toward them, so should IP law permit their imposition and provide remedies for their breach. This chapter shows that, with the exception of certain instances, this trend is misguided and should be resisted, not because the insights from modern antitrust are irrelevant, but because insights from modern antitrust do not support the case against the first sale doctrine. The main benefits of post-sale restrictions involve situations of imperfect vertical integration between co-producing or collaborating firms, which occur during the production and distribution phases or shortly thereafter. In such situations, contracting around the first sale doctrine should be permitted, and agreements containing such restraints should not be automatically condemned. Beyond such limited circumstances, however, the first sale doctrine promotes important social and economic goals: it promotes efficient use of goods embodying IP, guarantees their preservation, and facilitates user innovation, while minimizing transaction costs that otherwise might impede those goals. A closer look at what the insights from modern law and economics can teach reveals that rather than undermining the first sale doctrine they confirm its validity and support its continued vitality.