WHAT ANTITRUST LAW CAN (AND CANNOT) TEACH ABOUT THE FIRST SALE DOCTRINE

Ariel Katz*

ABSTRACT

The first sale doctrine (or exhaustion) limits the exclusive rights that survive the initial authorized sale of an item protected by such rights. The first sale doctrine has always been under pressure by owners of intellectual property rights, and courts have never been able to precisely outline its contours, or fully articulate its rationale. Recently, and somewhat counter-intuitively, insights borrowed from modern antitrust law and economics are invoked to provide a seemingly robust theoretical foundation for undermining exhaustion rules or narrowing their scope, and thereby strengthen IP owners’ control over downstream distribution and use of the goods they produce.

This article shows why this trend is misguided and argues that it should be resisted, not because the insights from modern antitrust are irrelevant, quite the contrary. Indeed, the insights from modern antitrust law and economics can help underpin some of the first sale doctrine’s missing theoretical foundations and help drawing its proper contours. However, as this article demonstrates, the insights from modern antitrust do not support the case against the first sale doctrine. When taken seriously, these insights ultimately support its continued vitality.

I. Introduction.............................................................................................................................2
II. The Origins...............................................................................................................................5
   A. The Antitrust Legacy of Bobbs-Merrill .................................................................7
III. Modern Antitrust Approach towards Post-Sale Restraints ..............................11
IV. The First Sale Doctrine: Different Versions............................................................14
V. The Flaw in the Parallel Trade Debate........................................................................17
   A. Price Discrimination ........................................................................................................18

* Associate Professor, Faculty of Law, University of Toronto. Director, Centre for Innovation Law and Policy. This article is based on a paper given at the Exhaustion and First Sale in IP Symposium, High Tech Law Institute, Santa Clara Law School, Nov. 5, 2010. I wish to thank Eric Goldman, Mark Lemley, Mark McKenna, Anne Layne-Farrar, and other symposium participants for their valuable comments and insights, as well as to thank my colleagues Peggy Radin, Abraham Drassinower, Ed Iacobucci and Michael Trebilcock for their advice. Special thanks go to Michael Laskey and Nicholas Van Exan for their superb research assistance.
WHAT ANTITRUST CAN TEACH

I. INTRODUCTION

Preparing for the new academic year, you find that some of the textbooks you are required to buy are available more cheaply abroad—can you import them, or do you need the copyright owners’ permission? You recently installed the latest version of your preferred word processing software—are you allowed to resell the copy of the older version? You are an electronics geek, enjoying disassembling old computers and building new devices from their components—do you need to get permission from the owners of the numerous patented inventions embodied in these components? The answers to these questions depend, in large part, on the scope of the first-sale doctrine.

The first sale doctrine in copyright law, or exhaustion rules of other intellectual property rights, limit intellectual property owners’ power to control the downstream distribution and use of products or copies embodying their inventions or their works, or bearing their trademarks.
When such rules apply, the patent owner’s right to sell a product embodying the invention, the copyright owner’s right to distribute copies of her work, or the trademark owner’s right to sell products bearing the trademark, are terminated after the first authorized sale of the genuine product or work. As a result, any further sale, rental or other disposition of ownership or possession (and in the case of patents, use of the patented invention) does not require the authorization of the owner of the intellectual property right.

Exhaustion rules are under pressure. They have always been. Understandably, intellectual property owners have never been great fans of such rules. While they do not always wish to control tightly the distribution of their products or copies, in some cases they might. All things equal, IP owners would rather have the legal power to exercise such control, and many of them argue against exhaustion rules and have tried working around them by various legal or technological means. So far, these attempts have only been partially successful. In copyright law, this partial success is reflected in ambiguous statutory language with regard to whether or not the doctrine applies to copies made outside the US, with language that ostensibly limits its application to copies that are owned by the downstream user, rather than licensed, and with uncertainty with regard to how freely transacting owners and users can contract around the doctrine, and the effect of such contracts on third party purchasers.

In patent law, even though the statutory language is simpler, court still grapple with defining

---

5 In addition to making, offering for sale and selling the patented invention, the Patent Act grants the patentee an exclusive right to “use” the patented invention, 35 U.S.C. § 154 (2010). Copyrights or trademarks do not confer such general exclusive right to “use”.
6 Costco v. Omega, 541 F.3d 982 (9th Cir., 2008), affm’d by an equally divided court, 562 U. S. ___ (2010).
9 Exhaustion rules haven’t been codified. The Act unambiguously grants an exclusive right to sell, and exhaustion rules, developed by the courts, do not require grappling with the ambiguous language such as the language of the copyright act.
the exact scope of exhaustion. If there is one thing that all can agree on, is that consensus on exhaustion, or a set of coherent rules delineating the doctrine, are difficult to come by. Confusion and ambiguity abound; not only in the US, but on a global level too. This is old news.

The real news is that increasingly (and perhaps somewhat counter-intuitively) modern antitrust law is relied on to heighten the pressure on exhaustion. Not that antitrust law can directly abolish such rules; obviously, antitrust law is incapable of granting new intellectual property rights or expand existing ones. However, insights derived from modern antitrust law, and especially modern antitrust law’s growing tolerance towards vertical restraints, are invoked increasingly to provide a seemingly robust theoretical foundation for undermining exhaustion rules or narrowing their scope.

In this article I argue that we should resist this pressure, but not because the insights from modern antitrust are irrelevant, quite the contrary. Rather, these insights can help underpin the doctrine’s missing theoretical foundations and can help draw its proper contours. However, as this article shows, the insights from modern antitrust do not support the case against exhaustion. When taken seriously, these insights ultimately support its continued vitality, albeit in a more nuanced way.

In a nutshell, a closer look into the economic theory that has informed modern antitrust law’s growing tolerance towards vertical restraints reveals that the recognized benefits of most such restraints occur primarily: (1) in situations involving joint-production between firms that are not fully integrated along the production and distribution chain; and (2) immediately after the first authorized sale or shortly thereafter. In such situations, parties should be generally permitted to contract around exhaustion rules when such agreements promote efficient production and distribution. However, the law should be generally averse to attempts to impose post-sale restraints on parties that are not joint-producers (e.g., consumers), and attempts to impose long-term post-sale restraints.

For ease of proposition, this article focuses primarily on the first-sale

---

10 Quanta, 553 U.S. 617.
11 See e.g., Euro Excellence (Canada), Article 6, TRIPS; EU rules creating EU exhaustion, etc.
12 See e.g., The First Sale Doctrine, Intellectual Property Colloquium (2010), at http://www.ipcolloquium.com/Programs/12.html (discussing the first sale doctrine and comparing copyright’s approach to that of antitrust law).
doctrine in copyright law, but much of its analysis is equally applicable to other IP rights. The article is organized as follows. It very briefly traces the origin of first-sale doctrine in copyright law and identifies one immediate lesson that antitrust law can teach, namely that unexhausted IP rights can facilitate anti-competitive practices. It then describes the modern view of antitrust law of vertical restraints, and how this view can inform the debate about the first sale doctrine. The article then proceeds to identify three common flaws in the application of the insights derived from antitrust law and economics and finally demonstrates how the insights, combined with more recent contributions to innovation theory, actually support the continuing vitality of the doctrine. More specifically, the article shows that contrary to the view of some commentators economic theory provides very limited support for laws banning parallel imports (i.e., the unauthorized importation of genuine goods sold abroad with the authorization of the IP owner). It explains why even when post-sale restraints increase efficiency and ought to be enforceable, they should be enforced primarily as a matter of contract law, not on the basis of intellectual property entitlements. Further, the articles demonstrates that while exhaustion rules should be treated as default rules, they should be treated as de jure sticky defaults, meaning that contracting around the first-sale doctrine should be presumptively invalid. Courts should refuse to enforce license conditions or contract terms limiting the ability of the user to resell goods embodying IP rights, unless the IP owner can demonstrate that the restraint is necessary and superior to other means to achieve efficiency.

Let me turn to the origins of the first-sale doctrine.

II. THE ORIGINS

Bobbs-Merrill v. Straus14 is often considered as the grand entry of the first-sale doctrine into American copyright law, following a similar principle of exhaustion in patent law.15 But like many legal rules, its genesis

---

13 The analysis can be applied even more broadly to all restraints on alienation of property, whether or not IP rights are involved.


15 See Motion Picture Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502 (1917); United States v. Univis Lens Co., 316 U.S. 241 (1942); see also Katherine J. Strandburg, Users as Innovators: Implications for Patent Doctrine, 79 UNIVERSITY OF COLORADO L. REV. 467, 495 (2008) (“The first sale doctrine holds that a patentee’s rights are ‘exhausted’ when a patented product is sold, leaving the purchaser free to do with it as he or she wishes.”)
can be traced to more than a single source or event. As others have noted, the genesis of the first-sale doctrine could be understood as a confluence of different legal streams: the common law hostility to restraints on alienation, to servitudes in general and servitudes in chattels in particular; the emergence of antitrust and its contemporaneous hostility towards all restraints: horizontal and vertical; and an era in which IP rights were more modest in their scope\textsuperscript{16} and greater regard was given to their limited nature.\textsuperscript{17} And perhaps, these early cases were decided in a more formalist era when legal thinking and reasoning tended to appreciate the distinction between legal categories and were averse to blurring the lines between them.

Each of these strands has gone through substantive transformation over time. The common law's aversion to servitudes has waned; we have grown to become accustomed to expansive IP rights and many of our policies have been shaped by a misguided full appropriability paradigm;\textsuperscript{18} complex regulatory systems have been part of our modern state, and while preachers of free enterprise continue to bemoan excessive government control some of them express no concern about excessive corporate control of individuals. Lastly, antitrust law, influenced by antitrust scholarship, not only has relaxed its hostility toward vertical restraints, but in some sense has turned to embrace them enthusiastically. Therefore, it is possible that just as the confluence of different legal streams bred the first-sale doctrine, a new confluence of these streams, flowing in reverse direction, will lead to its demise. Antitrust law has played and may play pivotal roles in these processes.\textsuperscript{19}

\textsuperscript{16} See e.g., Neil Netanel, \textit{Copyright and a Democratic Civil Society}, 106 YALE L. J. 283, 294 (1996) (characterizing modern copyright law as “Bloated”); see also Lionel Bentley, \textit{R. v. The Author: From Death Penalty to Community Service}, 20\textsuperscript{th} Annual Horace S. Manges Lecture (Tuesday, April 10, 2007), 32 COLUM. J.L. & ARTS 1, 3 (2008) (describing the consensus among academic about the expansion of copyright).

\textsuperscript{17} Bobbs-Merrill, 210 U.S. at 346 (“The copyright statutes ought to be reasonably construed, with a view to effecting the purposes intended by Congress. They ought not to be unduly extended by judicial construction to include privileges not intended to be conferred, nor so narrowly construed as to deprive those entitled to their benefit of the rights Congress intended to grant”). See also Motion Pictures Patents Company v. Universal Film Manufacturing Company, 243 U.S. 502 (1917).


\textsuperscript{19} Herbert Hovenkamp, \textit{Post-Sale Restraints and Competitive Harm: The First Sale Doctrine in Perspective}, NYU ANNUAL SURVEY OF AMERICAN LAW 1, 17 (2010) (“Over
A. The Antitrust Legacy of Bobbs-Merrill

Bobbs-Merrill, decided in 1908, is often described in the antitrust literature as a resale price maintenance (RPM) case: an attempt by a manufacturer (publisher) to maintain the resale price of her product (book). Bobbs-Merrill was the publisher of a novel titled *The Castaway* by Hallie Ermine Rives.20 A notice printed on the book prohibited anyone from reselling it for less than $1, and threatened to treat any violation of this condition as copyright infringement. Macy’s, the department store, purchased copies of the book from a distributor and sold them for 89¢ each. At that time Macy’s was a discounter—think of it as a Wal-Mart of old—and Bobbs-Merrill sued for copyright infringement. The Supreme Court dismissed the action, relying on the principle of exhaustion borrowed from earlier patent cases, as well as on the general policy against restraints on alienation.

Strictly speaking, Bobbs-Merrill was not an antitrust case, because Bobbs-Merrill sued for copyright infringement and the Court’s decision focused on the scope of a copyright owner’s exclusive right to vend copies. However, conventional wisdom holds that the decision reflects the contemporaneous hostility of the Court to all post-sale restraints. Three years after Bobbs-Merrill, the Court reiterated this hostility, this time as a matter of antitrust law, when it held, in *Dr. Miles v. John D. Park*,21 that resale price maintenance was illegal per se under the antitrust laws.22 According to Hovenkamp, this trend culminated in a duo of cases, *Straus v. Victor Talking Machine* and the *Motion Pictures Patents* case, issued on the same day in April, 1917, which effectively created a merger between the first sale doctrine and antitrust policy, and further planted the seeds of what would later become the related doctrine of intellectual property misuse.23

Viewed from this perspective, namely, that the first-sale doctrine is the IP mirror image of antitrust law’s per se prohibitions on RPM and tying, it is tempting to view antitrust law’s shift away from such per se rules as a

---


22 *Id.*, at 408, *overruled* by Leegin Creative Leather Prods., Inc. v. PSKS, Inc. 551 U.S. 877 (2007) (holding that resale price maintenance is subject to rule of reason analysis rather than per se illegality).

basis for similar leniency in IP law, in the form of a weakened (or even repealed) first-sale doctrine. After all, one might argue, if modern antitrust law is no longer hostile to post-sale restraints, IP rules rooted in the same hostility should be regarded as equally anachronistic. In the rest of this article I will explain why such an argument would over-simplify modern antitrust law and its bearing on the first-sale doctrine, but before this, I wish to highlight a less familiar point about Bobbs-Merrill’s antitrust legacy, as this hidden point can teach an important lesson as we proceed to discussing the first-sale doctrine’s future.

The truth is that Bobbs-Merrill wasn’t just an RPM case. The Supreme Court’s decision ends with a remark explaining that in light of the Court’s conclusion dismissing the copyright infringement claim, “it [is] unnecessary to discuss other questions noticed in the opinion in the circuit court of appeals, or to examine into the validity of the publisher's agreements, alleged to be in violation of the acts to restrain combinations creating a monopoly or directly tending to the restraint of trade.” This statement is the Court’s single allusion to antitrust, but it hints at a much richer antitrust history, which the opinion of the District Court below actually reveals.

It turns out that Bobbs-Merrill was not about a publisher preferring, for whatever reason, to exercise control over the distribution system of its books. What the case was truly about was a double-layered industry-wide cartel, established in 1900 under which the book publishers and the booksellers conspired to “correct some of the evils connected with the cutting of prices on copyright books” by requiring all publishers to fix the retail prices of their respective books, prevent discounting and eliminate discounters by collectively refusing to deal with them. RPM was just the means to enforce the cartel, and copyright was just an effective method to enforce the RPM.

In fact, litigation started in 1903 when Macy’s sued the American Publishers’ Association and the American Booksellers’ Association. Relying on the Sherman Act and its equivalent New York Anti-Monopoly Act, Macy’s attempted to enjoin the cartel. Macy’s bid, however, was only partially successful. The New York Court of Appeals found in favor of Macy’s on the basis that the publishers’ and booksellers’ rules sought to

---

prevent the discounting of all books, copyrighted and non-copyrighted alike.  

Relying on Bement v. National Harrow, however, the court held that a combination to fix the prices of copyrighted books alone would be legal. Soon thereafter, the respective associations’ rules were amended to cover copyrighted books only, and Bobbs-Merrill commenced its copyright infringement suit against Macy’s. In its defence, Macy’s argued that the notice did not restrict the right of an owner of a lawfully printed and sold copy of a book to sell it as it saw fit, and further defended on the grounds that the lawsuit was an attempt to enforce an unlawful combination. The Circuit Court sided with Macy’s on both issues, and refused to construe Bement v. National Harrow as broadly as the state court did. On appeal, the Second Circuit affirmed, but addressed only the copyright question. The Supreme Court, as noted above, mentioned the antitrust issue only in passing and based its decision solely on copyright grounds.

Nevertheless, this antitrust history of Bobbs-Merrill can teach an important lesson about the merit of the first-sale doctrine and the perils of abandoning it. It is well known that while cartels can be highly beneficial to their members (and detrimental to consumers) they are unstable because each of their members can do even better by cheating. Therefore, in order to be sustainable, cartel members need to be able to monitor cheating and enforce adherence to the cartel rules. RPM can be used as a means to enforce cartels. Retailers can use their collective power vis-à-vis a manufacturer and have the manufacturer “impose” RPM. If retail prices are fixed, cheating may be easily monitored, and the colluding retailers may demand that the deviant retailer be disciplined. RPM can also sustain a

---

27 Id, at 477-81.

28 E. Bement & Sons v. National Harrow Co., 186 U.S. 70, 91 (1902) (“[t]he general rule is absolute freedom in the use or sale of rights under the patent laws of the United States. The very object of these laws is monopoly, and the rule is, with few exceptions, that any conditions which are not in their very nature illegal with regard to this kind of property, imposed by the patentee and agreed to by the licensee for the right to manufacture an article, will be upheld by the courts. The fact that the conditions in the contracts keep up the monopoly or fix prices does not render them illegal.”)

29 Bobbs-Merrill Co. v. Straus, 210 U.S. at 172.

30 Id., at 156.

31 Bobbs-Merrill Co. v. Straus, 147 F. 15 (2nd Cir. 1906).

manufacturers’ cartel, by reducing individual manufacturers’ incentive to cheat and by providing means to monitor cheating.33

While cartels may employ contractual RPM to stabilize or enforce their rules, the utility of RPM is limited because it depends on the ability to enforce the contractual obligation, and the obligation will not be binding on third parties who purchased the goods from a cheating member and resell them at a lower price. In such situations IP rights, if unexhausted, can prove much more useful (hence dangerous). *Bobbs-Merrill* is a case in point. Despite being blacklisted for being a price-cutter, Macy’s had been able to procure copies of the book and offer them for sale at a lower price. This, of course, could destabilize the cartel because other booksellers would have to face the competition and would have to lower their prices too. Macy’s, however, was not bound by any RPM agreement and the Supreme Court holding guaranteed that it could, at least as a matter of law, continue to cut prices and destabilize the cartel. Had the Court ruled otherwise, that is, had the publisher been able to enjoin Macy’s from selling at a lower price on pain of copyright infringement, maintaining the cartel would have been much easier.

It could be argued that this hardly provides a justification for the first-sale doctrine, because if the problem is an underlying cartel, antitrust law will have no problem condemning it. The answer, however, is that the outright condemnation of cartels is only as good as there is evidence for its existence, and furthermore, tacit collusion, whereby competitors in an oligopolistic market can, without an agreement, lawfully coordinate their behavior to achieve a cartel-like outcome can also be bolstered by IP-based RPM. The fact that many IP based markets comprise a small number of large competitors,34 and that virtually any product can be sold accompanied by a trademarked and often copyrighted logo,35 should give rise to a concern that unexhausted IP rights can facilitate tacit collusion in such markets.

The full antitrust history of *Bobbs-Merrill* should make us wary of unexhausted IP rights, and we can summarize this lesson in the following proposition:

33 Id. at 329.


**Proposition 1:** **beware of unexhausted IP rights!**

The antitrust history of *Bobbs-Merrill* teaches an immediate lesson about the perils of unexhausted IP rights, namely their better capacity than contracts to facilitate cartels and other types of anti-competitive behavior. A broader lesson, which will be further developed as the article proceeds, is that attention should be given to the fact that post-sale restrictions enforced by rules of property may lead to many unintended consequences. Therefore, recognizing that post-sale restrictions may be efficient does not necessarily support harnessing property rules to enforce them. Careful attention should be given to the question of what legal instruments should be used to enforce them: contracts, property, or perhaps other rules.

**III. MODERN ANTITRUST APPROACH TOWARDS POST-SALE RESTRAINTS**

The law concerning post-sale restraints (as a subset of a broader type of vertical restraints) has changed dramatically throughout the history of antitrust law.

Section 1 of the Sherman Act declares illegal and punishable “[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations”.36 It is common ground today that although “every contract, etc.” may be subject to scrutiny, not all contracts are equal. One fundamental distinction is between horizontal agreements—those entered between actual or potential competitors, i.e., between firms at the same level of the distribution chain, and those entered between firms at different levels of the distribution chain (e.g., manufacturer-wholesaler; wholesaler-retailer; retailer-customer). Another distinction is between restraints that are deemed per se illegal and those whose illegality required proof of anti-competitive effect under a rule of reason. “Resort to per se rules is confined to restraints … ‘that would always or almost always tend to restrict competition and decrease output.’” … To justify a per se prohibition a restraint must have “manifestly anticompetitive” effects, .. and “lack … any redeeming virtue.”37

Early antitrust law was as averse to vertical restraints as it was hostile to horizontal restraints, and treated many agreements imposing vertical restraints, such as exclusive dealing, tying, and resale price maintenance as

---


37 *Leegin*, at 886 (internal citations omitted).
per se illegal. Gradually, however, antitrust scholarship began recognizing that horizontal and vertical restraints ought to be treated differently, because while vertical restraints might, at times, be used to achieve anti-competitive ends, in many cases they can and are used to enhance efficiency, prevent opportunism or otherwise advance pro-competitive outcomes. Jurisprudence followed scholarship, and since the late 1960, the Supreme Court gradually abolished virtually all per se rules applying to vertical restraints, with the last bastion falling in 2007 in *Leegin*, which overruled an almost century-old per se prohibition on resale price maintenance. While all such restraints could be found illegal under the rule of reason, abolishing per se prohibitions reflects the prevailing wisdom that such restraints are not necessarily harmful (according to some) or even almost always beneficial (according to others).

Without repeating the voluminous literature on vertical restraints, in order to understand why antitrust law has grown to tolerate vertical restraints but remained more hostile to horizontal restraints it may be useful to describe briefly the reason why horizontal restraints are perceived as more harmful than vertical restraints. The main concern about horizontal agreements is the elimination of competition between rival firms and the resulting ability to limit output, raise prices and collectively earn monopoly profits. The resulting harm is not limited to the higher prices that consumers have to pay and the resulting transfer of surplus from consumers to producers. Rather, the higher prices and the concurrent lower output means that some consumers who are willing to pay a price that covers the cost of production are forced out of the market, meaning that socially-efficient transactions are forgone. In the case of anti-competitive horizontal agreements the interests of the parties to the agreements are aligned but are diametrically opposed to the interests of consumers and society as a whole; the colluding competitors benefit from the reduced competition, while the costs are borne on consumers and society. In contrast, vertically situated parties do not share a common interest in reduced competition, quite the contrary. A manufacturer benefits when competition among its suppliers is intense, as well as when neither of its distributors possesses market power. Similarly, every distributor would rather be free to deal with competing manufacturers upstream and competing retailers downstream. It follows that this divergence of interests serves as a check on anti-competitive

---

38 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 ARIZONA L. REV. 837, 896 (2007).
vertical restraints. This basic insight leads to the following insights. First, a firm will not agree to the imposition of vertical restraints unless it is being compensated for the harm resulting from the restricted competition up-or downstream, and as a result the need to compensate makes anti-competitive schemes less attractive to begin with. Alternatively, the firm would agree to be bound by a vertical restraint because the restraint is actually efficient, in the sense that it allows the parties to increase output and share the increased gains from trade.

While it is always possible that vertical restraints can be anti-competitive, a growing body of literature has identified myriad reasons supporting the proposition that they often serve pro-competitive goals. For example, vertical restraints may encourage dealers to invest in developing a local market or in the supply of pre or post-sale services (such as training or repairs). I will discuss some of these pro-competitive explanations in Part V below.

The first-sale doctrine is relevant to post-sale restraints, a subset of the more general category of vertical restraints. Post-sale restraints often restrict what a buyer can do with the goods she purchased: where she can resell them, to whom, at what prices, whether she will have to provide pre- or post-sale services, repairs, warranties, etc. In the case of goods in which no IP rights subsist, contract law and the mere threat of termination serve as the main tool for enforcing such restraints. But generally speaking, remedies for breach of contract are weaker than remedies available for infringement of an intellectual property right. Remedies for breach of contract would probably be limited to expectation damages, and generally would be filed in state courts. Even though eBay v. MercExchange made the grant of injunctive relief less certain than before, injunctions are still more likely to be granted in IP infringement cases than for breach of contract, and the threat of injunction (especially when the user faces high

---

39 As Hovenkamp, I use the term “post-sale” restraint to refer “generically to any restriction imposed by a seller on how a purchased good can be used or resold after the initial sale” and in this context the term “sale” includes leases, licenses, or other transfers of interest short of a technical sale”, see Hovenkamp, supra note 19, at 1.

40 Contract law is the main, but not the exclusive tool. On some occasions plaintiffs were able to enforce the restraints against third parties relying on some common law commercial torts or even concepts of equitable servitudes created by notice, see e.g., Clairol, Inc. v. Sarann Co., 37 Pa. D. & C.2d 433, 1965 WL 6045 (Pa.Com.Pl.).

41 Hovenkamp, supra note 19, at 53.

switching costs),

Moreover, contractual restrictions are enforceable only against the purchaser who is party to the contract, but not against third parties such as the purchaser’s customers, users, or service personnel. Therefore, if IP rights subsist in the goods, producers can have a more powerful, and arguably more effective, tools to enforce post-sale restraints. However, if the IP rights are exhausted after the first sale, those more efficacious tools are no longer available.

IV. THE FIRST SALE DOCTRINE: DIFFERENT VERSIONS

From a modern antitrust perspective, as just described, it may be tempting to conclude that the first-sale doctrine is an inefficient IP rule, a relic from an era during which the benefits of vertical restraints were misunderstood, both in IP and in antitrust. After all, if post-sale restrictions are often efficient, then they ought to be enforced; and if they ought to be enforced, then why not provide the most effective tools for enforcing them? As Herbert Hovenkamp asks: “[does] the first sale doctrine serve[] any useful purpose at all?”

Before this question can be answered, it may be useful to note that the scope of the first-sale doctrine, despite more than a hundred years of litigation, is far from being clear. One unresolved issue is whether exhaustion should be national or international. That is, whether only a domestic first authorized sale exhausts the IP right, or any authorized first sale in any country. In addition, there could be several formulations of the doctrine, over a spectrum of strength. It is still somewhat an open question whether the first-sale doctrine is merely a default rule or a mandatory one, and if not mandatory, how sticky the default rule is. Different answers to

---


44 17 U.S.C. § 504(c).


46 Hovenkamp, supra note 19, at 54.

47 Id.
these questions can result in different formulations of the doctrine. Consider the following alternatives:

a) No First-Sale Doctrine

The spectrum of possible formulations can start with a no first-sale doctrine at all. Under this option, the first-authorized sale (or any subsequent sale) does not exhaust the IP right, unless the IP owner chooses otherwise. I include this option on the spectrum not merely for the elegance of the model, but because in some instances this is (or may be) the law. For example, if, as the Ninth Circuit held in Vernor, the doctrine has no application when a software publisher chooses to characterize a transaction as a “license” rather than a “sale”, then every subsequent sale or disposition of the copy requires the copyright owner’s permission. Similarly, if the first-sale doctrine applies only to copies made in the US, but not to copies made abroad, then possibly, the distribution right in such copies is never exhausted. In such situations, the IP owner can effectively dictate whether the doctrine applies or not.

b) Weak First-Sale Doctrine

Alternatively, the first-sale doctrine may be treated as a simple default rule. Under this version the IP owner’s exclusive right is exhausted only after the first unconditional sale, but it may not be exhausted if the transaction is conditional and conditions were imposed by license, contract or 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, subject only to antitrust scrutiny or antitrust-based IP misuse defence, meaning that the burden of showing an anti-competitive effect is on the user.

c) Moderate First-Sale Doctrine

Under a moderate version, the sale exhausts the IP rights, although the buyer may still be bound contractually (subject to antitrust/misuse scrutiny). That is, contracting around the doctrine may be permitted, but in case of

---

48 Quality King Distribs., Inc. v. L’anza Res. Int'l, Inc., 523 U.S. 135, 154 (1998), Ginsburg J., concurring; see also BMG Music v. Perez, 952 F.2d 318 (9th Cir.1991); Quality King v. L’anza, 523 U.S. at 154

49 But see Parfums Givenchy, Inc. v. Drug Emporium, Inc., 38 F.3d 477, 481 (9th Cir.1994) (holding that while the first-sale doctrine does not apply to copies made outside the US, it does apply upon the first authorized sale of those copies in the US).

50 This seems to be the view that Hovenkamp supports, supra note 19, at …
breach only contract remedies, not IP remedies, will be available.\footnote{This was the position of the United States in its Amicus Brief in Quanta Computer, Inc. v. LG Elecs., 553 U.S. 617 (2008), see Brief for the United States as Amicus Curiae Supporting Petitioners, 2007 WL 3353102 at 29 (“Some of the same restrictions that the first-sale doctrine renders ineffective in a patent-infringement suit could be validly imposed as a matter of state contract law. But even otherwise valid contract provisions would not provide a defense to a federal antitrust action, and mere unilateral notice to downstream purchasers will not generally give rise to enforceable contractual restrictions.”)}

d) Strong First-Sale Doctrine

A strong version treats the first-sale doctrine as a sticky default rule \textit{de jure}.\footnote{On sticky default rules, see generally Cass R. Sunstein, \textit{Switching the Default Rule}, 77 N.Y.U. L. REV. 106 (2002). While default rules often tend to be sticky, for the discussed in Part __ below, the first-sale doctrine may not be a sticky default unless its stickiness is bolstered legally.} That means that the sale exhausts the IP rights, and attempts to work-around exhaustion would be invalidated, unless the IP owner can provide a compelling case-specific explanation of why the work-around should be upheld.

e) Strongest First-Sale Doctrine

Under the strongest version, the sale exhausts the right, and attempts to work around it would be preempted, held invalid, or even constitute IP misuse \textit{per se}, i.e., without inquiring whether working around the doctrine can be justified.

Many antitrust-minded discussions about the first-sale doctrine range between advocating a weak version of the first-sale doctrine, or even no first-sale doctrine (at least in the context of parallel imports), to a moderate version. In this article, I posit that the optimal first-sale rule is actually a strong one, meaning that attempts to circumvent the doctrine would be presumptively invalid, unless the IP owner can provide a compelling justification for upholding them. In the following parts I will show that while the insights of modern antitrust justifiably reject the strongest form of the first-sale doctrine, relying on them for justifying the no/weak/moderate versions involves several flaws. The subsequent parts describe three such flaws. The first flaw arises in the context of the parallel trade debate, where proponents of IP rules preventing parallel trade increasingly rely on insights from modern antitrust. The second flaw occurs when the virtues of post-sales restrictions are invoked to justify a weak version of the first-sale doctrine, while failing to distinguish between the subtle, yet critical, difference between remedies for breach of contract and remedies for IP...
infringement, or whole actually embracing the notion of enhanced IP remedies, without recognizing their associated costs. The third flaw occurs when the first-sale doctrine is treated merely as a default rule, and assumes that contracting around it presumptively efficient. This approach is flawed because it adopts a Coasian logic in a notoriously non-Coasian setting, that of intellectual property. Let me turn now to a discussion the first of these flaws.

V. THE FLAW IN THE PARALLEL TRADE DEBATE

Debates around whether parallel trade should be permitted or banned increasingly rely on insights from modern antitrust law and economics. The term parallel trade refers to situations in which goods, produced by a manufacturer and sold abroad at a lower price than that charged locally, are imported (or re-imported) and sold domestically in competition with the same goods sold through the local authorized distribution system designated by the manufacturer. When such goods embed some protected IP, IP laws may be invoked to ban the unauthorized importation. Whether or not IP law should be used for this purpose is subject to a heated debate and the actual rules are often inconsistent among nations, between the different IP laws within a nation (i.e., different rules for patents, copyrights, and trademarks), and even within a specific national IP law. One major axis around this debate revolves is whether the IP right is exhausted only upon the first domestic sale authorized by the IP owner (national exhaustion), or whether any first sale authorized by the IP owner abroad exhausts the distribution (or importation) right (international exhaustion) as well. Under a third, hybrid, rule of “regional exhaustion”, any first authorized sale within a regional free-trade area would exhaust the IP rights, but IP laws

53 See ROSE ANN MCGILLIVRAY, PARALLEL IMPORTATION (2010).

54 See e.g., Euro Excellence (compare: Rothstein, J, holding that parallel imports of a copyrighted work infringes the copyright only when the local copyright holder is an assignee rather than an exclusive licensee; Abella, J., rejecting the distinction between an assignee and an exclusive licensee and holding that the unauthorized importation infringes copyright in both cases; and Bastarache, J., agreeing with Abella that the distinction between an assignee and an exclusive licensee is immaterial, but arguing that the ban on importation exists only when the work is imported as a work, but not when the work is a logo printed on a wrapper of a consumer good.), for a thorough discussion of the different opinions see ABRAHAM DRASSINOWER, THE ART OF SELLING CHOCOLATE: REMARKS ON COPYRIGHT’S DOMAIN, IN FROM "RADICAL EXTREMISM" TO "BALANCED COPYRIGHT": CANADIAN COPYRIGHT AND THE DIGITAL AGENDA (MICHAEL GEIST, ED., IRWIN LAW, 2010).
might still be used to block unauthorized imports from all other countries.\textsuperscript{55} Regional exhaustion is currently the law in the European Economic Area (consisting of the European Union and Iceland, Lichtenstein and Norway).\textsuperscript{56}

Supporters of international exhaustion usually highlight the benefits of increased competition and lower prices that parallel imports can bring about,\textsuperscript{57} while opponents of such rules often emphasize the benefits arising from tight control by producers over their distribution systems.\textsuperscript{58} They point out that parallel imports undermine the integrity of such systems, resulting in less efficient distribution systems, to the detriment of producers and consumers alike. Consequently, they resist international exhaustion and support adherence to domestic exhaustion.

The argument supporting banning parallel imports will be familiar to anyone acquainted with modern antitrust thinking about vertical restraints. The arguments typically include one or more of the advantages, discussed below, which would arguably materialize in a system that bans parallel imports, but would be stymied if parallel trade were allowed. This Part discusses these arguments and shows that they fail to support an IP-backed ban on parallel imports, namely national exhaustion. As I explain below, these arguments can lead to only two sensible rules: either international exhaustion or no exhaustion, but they cannot sensibly justify national exhaustion.

\textit{A. Price Discrimination}

Proponents of national exhaustion often invoke international price discrimination as a justification for banning parallel trade. Price discrimination may occur when demand for a product varies across countries and producers might be interested in setting different prices in such countries: to charge a higher price where demand is higher and lower where demand is lower. Parallel trade is a form of arbitrage, which


\textsuperscript{58} Infra.
jeopardizes the producers’ ability to price-discriminate. Proponents of national exhaustion often argue that price-discrimination promotes consumer welfare and therefore should be encouraged, and consequently, that parallel trade should be prohibited. Price discrimination is considered beneficial for the following main reasons: (1) it increases output compared to a situation in which a firm with market power can only set uniform prices, thereby ameliorating some of the inefficiencies otherwise associated with the exclusivity inherent in IP rights. The ability to charge different prices will actually benefit consumers with lower ability to pay (e.g., in developing countries) that would not be able to obtain the product if the seller could only set a uniform price (fearing that the low-priced products will be exported to high-priced countries); (2) the increased output may contribute to achieving economies of scale and learning resulting in lower per-unit costs; and (3) even when price discrimination does not increase output, the additional profit will encourage investment in R&D and innovation and contribute to dynamic efficiency.

In sum, price discrimination should be encouraged, because it promotes allocative efficiency (i.e., leads to higher output and/or lower costs) as well as dynamic efficiency (encourages innovation). But opponents of international exhaustion maintain that price discrimination is not only efficient but also just (because otherwise consumers with low willingness or ability to pay will not be served). To the extent that the first sale doctrine put sand in the wheels of price discrimination, it leads not only to greater deadweight loss, but also to distributional injustices.

No other area demonstrates these points more dramatically than pharmaceuticals, where the affordability of drugs can be a matter of life or death. It is often assumed that the high cost of developing new drugs requires patent protection and justifies the high prices charged for those

59 See e.g., David A. Maluég & Marius Schwartz, Parallel Imports, Demand Dispersion, and International Price Discrimination, 37 J. INTL. ECON. 167 (1994).

60 Id, at 175-181 and 189-190; contra, see Frederick M. Abbott, First Report (Final) to the Committee on International Trade Law of the International Law Association on the Subject of Parallel Importation, 1 J. INTL. ECON. L. 607 at 619-621.

61 Maluég & Schwartz, supra note 59, at 190.


drugs. Yet those drugs are beyond reach for millions of people in developing countries. Price discrimination seems like the obvious solution: drug companies would set high prices in the rich developed countries, and lower prices in less-developed one, and will be able to spread the cost of developing them across all markets, as well as make the drugs available throughout.\(^{64}\) It is often argued that drug companies would readily do that if they could prevent the cheap drugs from being re-imported into developed countries’ markets, but if they cannot, they would rather not sell in the less-developed countries at all.\(^ {65}\) Banning parallel trade, therefore, seems like a just and efficient rule. To increase the dramatic effect, we could say that the first sale doctrine kills.

While trade in pharmaceuticals vividly highlights the issue, it can be a red herring.\(^ {66}\) First, trade in pharmaceuticals is strictly regulated in most countries and in all developed countries. Even if an astute arbitrageur identified an opportunity to import cheap drugs from a low-price country, pharmaceutical regulation laws would not allow that or at least make compliance costs prohibitively high, irrespective of the importing country’s exhaustion regime. One could then assume that differential pricing would be prevalent, and drugs would be universally affordable. This is not the case.\(^ {67}\) The fact that drugs are often still unaffordable in many developing countries despite the illegality in parallel trade suggests that there are other reasons for the lack of widespread price discrimination. One reason is that many developing countries have a small population of highly affluent citizens and a majority of extremely poor citizens. In such cases, the local profit-maximizing price in the developing country is at a developed country


\(^{65}\) See e.g., Alan O. Sykes, *TRIPs, Pharmaceuticals, Developing Countries, and the Doha “Solution”*, 3 CHI. J. INT’L L. 47, 64 (2002).

\(^{66}\) The EU is an exception. As part of the objective of creating a common market, the EU adopted a regional exhaustion regime. As a result, IP rights cannot be used to prevent the trade in pharmaceutical in between the member states. The EU has also harmonized drug regulation with the effect that pharmaceutical regulations laws cannot be used to prevent intra-EU parallel trade. However, price controls of pharmaceuticals remain within the jurisdiction of each member state. As a result, parallel traders take advantage of the differentials in price-controls to import drugs from lower regulated prices into those countries with higher regulated prices.

level.\textsuperscript{68} Another reason might be that pharmaceutical companies may refrain from charging lower prices in developing countries because they may be concerned about populist backlash in developed countries if legal versions were available cheaply abroad, which may result in the introduction of price controls (in the US), or tighter price controls (in many other developed countries), even if parallel importation of drugs is prohibited.\textsuperscript{69} Lastly, even if a compelling case can be made that parallel trade in pharmaceuticals will prevent efficient price discrimination, the fact that pharmaceutical trade is already highly regulated makes it easy to regulate such trade as part of this regulatory framework, without hijacking the entire area of IP for this purpose.

Second, it is well known, even if sometimes ignored, that whether price discrimination results in greater or lower output compared to uniform pricing cannot be determined a priori. Depending on the characteristics of the individual market, a single monopoly price can be set at a lower or higher level of output compared to that under price discrimination. Therefore, it cannot be determined a priori whether international exhaustion harms or improves allocative efficiency.

Third, as Wendy Gordon\textsuperscript{70} and Michael Meurer\textsuperscript{71} have pointed out, even if price discrimination is more efficient than a monopoly uniform price, it is not more efficient than a competitive price. The first sale doctrine is one of those internal IP rules that mitigate the social cost resulting from the IP exclusivity. Every new generation of a product (e.g., an iPhone nG) competes to some extent with the products from the previous generation sold as second hand (e.g., iPhone (n-1)G). Although the competition is imperfect (as the products are not perfect substitutes), some potential consumers might choose a second hand product if the price of a new one is too high.\textsuperscript{72} In order to sell to them the manufacturer must either lower the

\textsuperscript{68} Id., at 930.

\textsuperscript{69} Id. Of course, they may also be concerned that developed countries will lift the prohibition and allow parallel imports from low-price developing countries, but the point is that even when parallel trade is not legally possible, pharmaceutical companies do not engage in the kind of international price discrimination that opponents of international exhaustion would expect them to engage in.


\textsuperscript{72} Acknowledging that used goods and new ones are substitutes does not, of course,
price, or improve the quality of the new version significantly enough to render the older one less attractive. When this is considered, it becomes less clear that allowing the producer to prevent resale leads to a more efficient outcome.

Of course, buying a second hand set is not ideal and if the manufacturer could price discriminate and sell all consumers an nG iPhone, both the manufacturer and the lower-demand consumers would be better off. Preventing such lower-demand consumers from buying an nG iPhone at a lower price and turning around to resell it to higher-demand consumers might be necessary to guarantee that they will get the nG iPhone at all. Arguably, then, greater ability to prevent resale could lead to the best outcome for all. However, even without the ability legally to prevent resale the producer could still price-discriminate by reducing the incentive to resell (for example, by voiding warranties upon resale, bundling the product with services such as airtime, etc.). But even if the inability to legally prevent resale prevents the supposedly more efficient outcome (nG iPhone for all), the need to compete with the (n-1)G iPhone might be the reason why the more attractive nG product is available at all. That is because a producer who can eliminate a second hand market faces lower incentives to innovate. If the producer does not face imminent inter-brand competition (e.g., an nG Blackberry), he can delay the development and offering of the nG product until all consumers have purchased the (n-1)G. In such situations, a legally enforceable post-sale restraint may slow down innovation.

Fourth, the pharmaceuticals example can be misleading because its lessons are not easily applicable to other settings. Patented drugs represent a case that is very close to a stylized model favored by economists. We can assume that on average, individuals’ preference for their health and life is identical across countries, and that differences in demand reflect only

tell us how close substitutes they are, see e.g., Anindya Ghose, Michael D. Smith & Rahul Telang, Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact (September 2005). Available at SSRN: http://ssrn.com/abstract=584401 (finding that only 16% of used book sales at Amazon cannibalize new book purchases. The remaining 84% of used book sales apparently would not have occurred at Amazon's new book prices, and estimating that the used book market creates net welfare gains that far outweigh the losses to publishers). See also Rahul Telang, Welfare Implications of Used Markets. (Telang’s article from this symposium).

73 This is a variation on Arrow’s argument about the monopolist’s lower incentive to innovate.

74 Individuals may have different preferences (some are more health conscious than
differences in affordability (that is, the only reason why a middle-class American is willing to pay for a life-saving drug and a middle-class South-African may not is not their different valuation of life but the different ability to pay). We can also assume that the market structure is similar across nations, in the sense that a patented will have the same type and number of therapeutic substitutes in all countries. If so, difference in demand across countries will not reflect differences in the cross-elasticity of demand between the drug and its substitutes within each country. In other words, variation in prices across countries will indeed reflect differences in ability to pay rather than differences in how competitive one market is compared to another.  

Next, the cost of developing new drugs and the truly global nature of the market makes the proposition that price discrimination will enable patentees better to spread the cost globally and maximize the recovery of their investment theoretically compelling, at least more compelling than the argument that international price discrimination is essential for the production of, say, a new sound recording. Lastly, a situation in which people might die or suffer because they cannot afford paying for an available drug offends very deep notions of morality. If patents make the iPhone more expensive, the fact that some consumers cannot afford it may be considered to be an unavoidable but acceptable outcome of our choice to encourage innovation by granting patents. When patent-related monopoly pricing creates not only deadweight loss but also dead people it is more difficult to be sanguine about it. Under such conditions it is not difficult to see how price discrimination might be desirable, and why, to the extent that parallel trade undermines it, it should be controlled.

However, even if pharmaceuticals present a plausible case for restricting parallel imports (at least in theory), this stylized model does not easily

---

75 That is, if the patented drug is the only therapeutic remedy it is a monopolist in all countries. If there are substitutes, they are available in all countries. This assumption might be too simplistic. Countries with small markets (either because they are small or because they only have a small number of residents who can afford paying) can be less competitive than larger markets because it may be more difficult to achieve economies of scale in distribution, marketing, regulatory compliance, etc. Prices in such cases may be higher.

76 Despite this theoretical support, many developing countries strongly support parallel importation of drugs, and perceive it as an important policy lever for increasing the affordability of medicines in their countries. This evidence suggests that pharmaceuticals companies do not price-discriminate as extensively as the theory would predict.
apply to other products. Differences in local demand may be the outcome of a myriad of reasons, with the result that it is impossible to generalize and determine a priori whether price discrimination should be encouraged or not. For example, price differences may also be a function of differences in the cross-elasticity of demand between the product and other products within each country. That is, consumers in countries A and B, with the same disposable income, might be willing to pay different prices for the same product because they have different preferences for the product, given available alternatives. For example, consumers in Ontario might be willing to pay a higher price for ice-skates than consumers in Florida, because they have fewer options for outdoor sports during the winter months (in economic terms, the cross-elasticity of demand between ice-skates and other sports equipment is high in Florida and low in Ontario). If trade is free and shipping costs are not too high, an ice-skates monopolist may not be able to price-discriminate by segregating the Ontario and Florida markets and may choose either to sell at a higher price in Ontario (and price consumers in Florida out of the market) or to set prices in both locations at the Florida level. We cannot determine a priori which outcome is more likely. Nor can we say a priori that maintaining the incentive to invest in developing ice-skates requires an ability to segregate the Ontario and the Florida markets.

Consider another possibility. Under the same assumption as before, it may turn out that because the Ontario market is large it attracts many ice-skates manufacturers and becomes very competitive. The Florida market, however, consists of a very small number of ice-skating devotees who are willing to pay a much higher price. Under those circumstances free trade will allow the Florida consumers to take advantage of the competition in Ontario. It is theoretically possible that the marginal loss from the inability to sell at a higher price in Florida would undermine the incentive to invest in ice-skates development, but we have no way of knowing that a priori (and it does not seem likely anyway).

Therefore, if we cannot tell with any degree of confidence that laws prohibiting the importation of skates from Florida to Ontario or vice versa are preferable to free trade, there is no rational basis for banning such trade.

Price discrimination can also reflect differences in preferences resulting from differences in information available to consumers. In a recent paper supporting legal bans on parallel trade and exalting the benefits of price discrimination, Nicolas Petit provides the example of a bottle of Polish-branded vodka that “might be perceived as a special product in Western Europe – hence western EU consumers are ready to pay a high price for this product – and by contrast, be perceived as a relatively standard product in
Poland – hence Polish consumers are only willing to pay a low price for this product. In a setting of this kind, the vodka producer will charge different prices in Poland and in the western European countries, according to the consumers’ ability to pay (economists talk of consumer elasticity).\footnote{Nicolas Petit, “The Economics of Parallel Trade – Iconoclast Views on a Dogma of EU Competition Law” (2010) at 3, citing A. Heimler, “Competition Law Enforcement and Intellectual Property Rights” (2007), mimeo.} Petit maintains that there is “nothing intrinsically bad” about charging different prices in this situation.\footnote{Id, citing National Economic Research Associates (NERA), SJ BERWIN & Co et IFF Research, The Economic Consequences of the Choice of a Regime of Exhaustion in the Area of Trademarks, Final Report for DGXV of the European Commission 40, (8 February 1998) at 5, available at http://ec.europa.eu/internal_market/indprop/docs/tm/report_en.pdf.} If what he means is that the producer should not be morally or legally condemned for the fact that some consumers are willing to pay a premium for his brand (and for pricing accordingly) I agree. But there is nothing particularly laudable about it either, and it hardly follows that adopting laws bolstering this outcome are justified—quite the contrary. If the better-informed Polish consumers are not willing to pay a premium for the particular brand, then we can assume that the price they are willing to pay is the more efficient one. Therefore, there is no reason why the less informed consumers should be legally prevented from being offered the same price. This is exactly one of the virtues of competitive markets, where prices are set according to the reservation price of the marginal consumer (which can be the one who is better informed).\footnote{See MICHAEL TREBILCOCK, THE LIMITS OF FREEDOM OF CONTRACT (Cambridge: Harvard University Press, 1993) [pinpoint needed].} Furthermore, it seems highly unlikely that the producer would choose to quit the Polish market to prevent his vodka from being exported from there,\footnote{The Polish home market could be the main (home) market; quitting this market could ruin the brand (if the reason why west EU consumers are willing to pay a premium is that they perceive it as the authentic Polish vodka.} and even if he did, other Polish producers would easily fill the demand.

In sum, a plausible theoretical case for legal rules supporting price-discrimination and preventing arbitrage can be made in the case of pharmaceuticals, to the extent that the local differences in demand reflect only differences in the ability to pay and if such rules will, in effect, lead to greater affordability and greater incentives to innovate. Whether these conditions exist or not is beyond the scope of this article, but even if they exist, it is not clear at all that these lessons from pharmaceuticals are easily
transferable to all other settings involving trade in goods embodying intellectual property, where the welfare implications of price discrimination, both static and dynamic, are highly ambiguous. Therefore, if such rules are desirable for pharmaceuticals they could be incorporated into the already existing elaborate system of drug regulation, but there is no need to subordinate the entire area of intellectual property law for that goal.

B. Other Efficiencies of Post-Sale Restraints

1. Parallel Trade and Pre- and Post-Sale Services

In addition to the perceived benefits of price discrimination, antitrust scholarship has identified a variety of other benefits arising from post-sale restraints. The marketing of almost every good requires some level of local investment in establishing a distribution network, complying with local regulations, creating demand through advertisement and other marketing efforts, offering pre-sale services for potential customers (such as demonstration, training, advice) as well as post-sale services (such as technical support, handling repairs and returns, respecting warranties, etc.).

A manufacturer that is fully integrated into distribution and retail can ensure, through managerial orders, that all of these tasks are fulfilled. However, full vertical integration may not be possible or even desirable, and the manufacturer may need to enter into contracts with local firms to provide these services. Obviously, local dealers (distributors or retailers) may be reluctant to make such investments if they are concerned that after incurring the costs, consumers, rather than buying from them, will buy from other dealers offering the goods at a lower price. To ensure such investments, the manufacturer may 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 ensure investment and those investment are indeed necessary to increase output, then, the argument goes, these restraints are efficient, because they will result in more units sold and better services delivered.

From this perspective, parallel trade can be seen a cause for concern, because those who profit from parallel trade are the foreign dealer and the parallel importer, who free ride on the investments of the local dealer. Therefore, the argument goes, parallel trade undermines the incentive to invest in building the local market and provide pre and post-sale services, ultimately to the detriment of the local dealer, local consumers, and the manufacturer.

2. Parallel Trade and its Effect on Positional Goods

Opponents of parallel trade sometimes argue that parallel trade and the
resultant availability of cheap products sold via unauthorized outlets may undermine the value or the prestige of brands, especially luxury brands, to the long term detriment of consumers.\footnote{NERA, supra note 67.} This view resembles recent literature on RPM, which identifies it “as a production technique of status goods” that is “socially superior … to other production methods for status goods.”\footnote{Barak Y. Orbach, “The Image Theory: RPM and the Allure of High Prices”, 55 \textsc{Antitrust Bull.} 277, 303 (2010).} Barak Orbach notes that “[i]f RPM serves the product image, banning its use may undermine markets for status goods.” He argues that whether or not this outcome is desirable, “antitrust laws are not intended to accomplish such goals.”\footnote{Id.}

In line with Orbach’s observation about antitrust, IP law should equally be agnostic to the sale of positional goods. The law should not necessarily ban them and should probably permit distribution methods to maintain them, but this hardly justifies adopting rules to encourage them.

\textit{C. The Irrelevance of Antitrust Insights to the Parallel Trade Debate}

From price discrimination to brand image, modern antitrust scholarship, shaped by modern economics, has identified a variety of benefits that post sale restraints promote, and contributed to the gradual erosion of antitrust law’s earlier hostility towards vertical restraints. These insights are often invoked to support bans on parallel imports and to justify adherence to national exhaustion rules and reject international exhaustion. This section explains why notwithstanding the importance of these insights, they have very limited bearing on the question of whether IP law should adopt a model of national, international, or regional exhaustion. In fact, these insights highlight that the only sensible choice is between international exhaustion and no exhaustion at all. National exhaustion and its cousin, regional exhaustion, are legal hybrids that for the most part lack any economic rationale.

The reason is quite simple. Consider price discrimination. Even if or when price discrimination is efficient and arbitrage is harmful and should be prevented by law, banning parallel trade supports only one type of price discrimination: cross-country price discrimination. But price discrimination along national lines is a very crude type of price discrimination. It makes sense only when consumer demand within national boundaries is
homogenous but demand across nations is heterogeneous. Generally, however, the preferences of consumers within a nation will be just as varied as the preferences of consumers across nations. Therefore, it is unclear why IP law should target only international arbitrage, but remain deferential to inter-state, inter-city, or inter-personal arbitrage. In fact, if the goal of IP law is to encourage price discrimination and prevent arbitrage, then the only sensible rule is no exhaustion at all.

Consider the ice skates example discussed earlier. Assume that in addition to Ontario and Florida, consumers in Michigan have the same demand for ice-skates as their neighbors in Ontario. Also assume that some ice skates are patented or that a copyrighted logo is embedded in them. If price discrimination is desirable, then the optimal pricing would be to set a higher price in Ontario and Michigan and a lower price in Florida, and if IP law were to guarantee this pricing structure, it should prohibit the importation of skates from Florida into either Michigan or Ontario or both. Note, however, that a national exhaustion rule will only prohibit trade between Florida and Ontario, but not between Florida and Michigan. Consumers in Michigan will be able to import cheaper skates from Florida if the prices in Michigan were too high, but consumers in Ontario will not. National exhaustion will help segregate the market, but on the basis of an economically irrelevant dimension: national borders, not differences in demand. If IP law should intervene to prevent arbitrage, then national exhaustion misses the mark. Not only does it segregate markets according to an economically irrelevant criterion, but also it targets international arbitrage, where higher natural barriers often exist, while neglecting domestic arbitrage, which often faces no such barriers. It operates where it is less necessary. Therefore, if exhaustion should be limited to facilitate price discrimination, then no exhaustion ought to be the rule (and to achieve internal consistency, IP owners should be granted not only an inexhaustible right to sell; buying without their permission should constitute an infringement as well. After all, consumers from Michigan or Ontario who

---

84 This conclusion does not exclude the possibility that national border do provide an economically sound basis for distinguishing between domestic and international arbitrage. If, for example, parallel trade is a response to exchange rate fluctuations, then a cross-border arbitrage may be more prevalent than domestic arbitrage (and arguably more harmful to the local distribution system). While there is evidence that parallel importers are sensitive to currency fluctuations, the significance of this single factor is less than clear, see MacGillivray, supra note 53, at 27. In any event, it is not clear that manufacturers cannot provide sufficient protection to their local dealers against parallel trade by resorting to contractual means, or that there are other effective ways to hedge the risks of currency fluctuations.
travel to Florida to do shopping are just as free riders as unauthorized dealers who facilitate the arbitrage).

Moreover, national exhaustion supports only one type of price discrimination, geographical price discrimination across national borders. It does not support other forms of price discrimination and the elaborate methods that sellers often employ to maintain them.

In sum, while the virtues of price discrimination are often invoked to justify IP-based legal ban on parallel trade and support national exhaustion rules, there is a serious mismatch between the symptom (arbitrage) and the remedy (national exhaustion). Should IP law be harnessed to prevent arbitrage, no exhaustion should be the rule. Alternatively, and preferably, IP law should simply stay out of the picture when international arbitrage is concerned just as it has no power to prevent domestic arbitrage. In fact, the ubiquity of price discrimination schemes that occur domestically, notwithstanding the first-sale doctrine, should discount the arguments highlighting the necessity to prevent arbitrage.

The same point can be made with regard to most of the other identified benefits of vertical restraints. It is indisputable that establishing and maintaining efficient distribution systems benefits producers and consumers alike, and that often this goal requires the imposition of enforceable post-sale restraints. However, if IP law should play a role in achieving such a goal, national exhaustion provides a very partial tool. A manufacturer might be interested in assigning exclusive territories to different dealers, or to appoint one dealer to deal with one type of customers and another dealer with another type. Each of these dealers might need to sink costs in making some specific investments and might be reluctant to do that without being offered a credible protection against free riding by other dealers or the manufacture. Arguably, the first-sale doctrine undermines the credibility of such guarantees, because at some point, third parties might be able to obtain the goods and undercut the local dealer. If exhaustion rules should be limited to prevent that, national exhaustion seems like an arbitrary and unprincipled choice. Therefore, the only principled choices are either no exhaustion at all, or, if one finds the arguments about the detrimental effects of the first-sale doctrine not compelling, international (or indeed universal) exhaustion.

This Part explained why the myriad virtues of post-sale restraints provide very little help in settling the debate about exhaustion rules in the context of parallel trade. It might be tempting, at this point, to conclude that the first-sale doctrine, by undermining the efficacy of post-sale restraints, is indeed an inefficient IP rule, a relic from an era in which the economics of vertical restraints were not well understood. The next Part suggests that
such a conclusion should not be arrived at precipitously. Recognizing that some post sale restrictions are efficient is a necessary condition for their enforceability, and antitrust law undoubtedly moved in the right direction when it stopped treating agreements imposing such restraints as per se unlawful. But it hardly follows that they should be enforced as a rule of property, let alone a default one.

VI. THE PROPERTY/CONTRACT FLAW

As noted, modern antitrust teaches us that post-sale restrictions are not necessarily harmful and that furthermore, many of them may actually be quite beneficial and necessary to organize sophisticated distribution systems, when a manufacturer is not fully integrated into distribution and retail. It may be tempting, therefore, to adopt the following line of reasoning: if post-sale are efficient, they should be enforceable; and if enforcing them on the grounds of IP infringement is easier than as breach of contract, an IP remedy should be available. This line of reasoning is tempting, yet generally flawed. Remedies for infringement of a property right may be more effective than those available for breach of contract, but whether greater efficacy is desirable depends not only on the benefits of more compliance but on the costs that may be externalized to third parties. 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.

Whether post-sale restraints are imposed to facilitate geographic price discrimination or to encourage investment in building a distribution system or in the provision of pre- or post-sale services, antitrust scholarship has focused on the relationships between collaborating firms attempting to organize an efficient production and distribution system. We can generalize this insight to the following proposition:

**Proposition 2:** When firms jointly participate in a productive enterprise that is prone to opportunism, various enforceable restrictions may be necessary.

Support for Proposition 2 can also be derived from the New Institutional Economics literature that explores the connection between the theory of the

---

firm and intellectual property rights. Several scholars have demonstrated the importance of intellectual property rights for overcoming the acute problems arising from the incompleteness of contracts when firms involved in joint production need to contract around information. The main insight is that information presents a unique challenge for efficient contracting. Authors of new works, or inventors of new inventions, will often lack other complementary assets necessary for their successful commercialization. Therefore, they will need to collaborate with other co-producers, and they will need to disclose their new information for this purpose. However, after A (author/inventor) has disclosed her new information to B (publisher/producer), the information cannot be untaught. This creates an acute problem. A knows that B, having learned the new information, may renege on whatever promise he gave, and therefore will demand to be paid in full in advance. B, not knowing what the information is and whether it is valuable at all, will not agree to pay until he learned what the information is. Even if the parties can ultimately draft a suitable contract, the contract will be inherently incomplete because drafting the contract without defining what the information is challenging, and in any event, the contract will not be binding on third parties who may have learned the information from B. Intellectual property rights facilitate knowledge-transfer and co-production, by creating an enforceable asset that is easier to contract about. This reduces the cost of “deterring opportunistic rent seeking by team members while simultaneously increasing trust and cooperation within the team.”

Proposition 2 is also consistent with Stephen Maurer and Suzanne Scotchmer’s “profit neutrality” principle, which they define as “the principle that a patentholder’s reward should not depend on whether he has the ability to work the patent efficiently himself.” Applied to the present context, “profit neutrality” means that IP owners should be able to impose post-sale restrictions to achieve the same profit that they would be able to

86 See e.g., David J. Teece, Capturing Value from Technological Innovation: Integration, Strategic Partnering, and Licensing Decisions, 18 INTERFACES 46 (1988).

87 If the information is tacit, and B will not learn it unless A spent the time and effort to teach it, B will be concerned that after receiving the payment, B may fail to perform its obligation to teach. See A. Arora, Licensing Tacit Knowledge: Intellectual Property Rights and the Market for Know-How, 4 ECON. OF INNOVATION AND NEW TECH. 41 (1995).


gain if they were fully integrated into distribution and retail. In such situations of joint-production treating IP rights as default entitlements and permitting contracting around them to achieve efficiency in a profit-neutral way makes sense.90

However, the teachings of the literature that supports Proposition 2 cannot immediately apply to restraints that do not involve situations of joint production, and cannot generally support the imposition of enforceable post-sale restraints against third parties. For example, even if and when territorial restrictions should justifiably protect a local dealer from opportunistic behavior by other members of the same distribution system (other dealers or the manufacturer), it does not directly follow that third parties, who are not members of the same distribution system, who lawfully purchased a product in another territory should be prohibited from reselling them in the local territory. The cost/benefit calculus of making the territorial restriction enforceable against third parties is different. The difference is not that restraining the third party cannot possibly be efficient—it obviously can. From a strict economic point of view, a high-demand consumer who avoids paying a high price under a price discrimination scheme and goes shopping in a low-price territory, or buys from a parallel importer, or a consumer who takes advantage of the pre-sale advice given at a local showroom and then purchases the good elsewhere, is just as much an arbitrageur or free rider as the dealer who sells into another dealer’s territory. However, the likelihood that sales made by third parties will seriously undermine the viability of the authorized distribution system probably diminishes the further removed is the third party from the authorized channel, and consequently, the need to control third parties diminishes too. At the same time, in order to be enforceable against third parties whose identity cannot be determined ex ante, such restraints should be in the form of an in rem, or property entitlement. But precisely because the restraints might be enforced in rem, it may be enforced even when and where the reasons for imposing it no longer exist or in situations where enforcing it may not be socially beneficial. Moreover, such an in rem right might interfere without good reason with buyers’ freedom to seek the best deal available, which drives competitive markets and innovation in the first place.

For example, consider a company introducing an innovative product

---

such as the iPad. Assume that successfully marketing such a product requires setting up a distribution network that involves at least some retailers who can display the new product, allow potential consumers to touch and feel it, and provide on-site demonstrations and training. Obviously, retailers providing such services face higher costs than retailers who do not, and they might be reluctant to provide them if consumers could take advantage of these services, but then purchase from another retailer (e.g., online) selling the product at discount. This is the paradigmatic case in which post-sale restraints, limiting the ability of one retailer from undercutting the other, might be justified. Such post-sale restraints could be imposed by contract, and if the product is desirable enough for the retailer the threat of termination might be enough to secure compliance. However, contractual restraints are only enforceable against the contractual parties, so they would not protect the retailer from competition from other sellers who are not bound by any contractual restriction. For example, suppose that the manufacturer has implemented a price discrimination scheme under which some dedicated retailers sell the product to some consumers at a significant discount.\footnote{E.g., Apple’s education pricing, see http://store.apple.com/us-hed/findyourschool} Such retailers might be bound contractually from competing with the retailers in the other segment, but these contracts will not bind third parties, such as arbitrageurs buying the product at the discounted price and then reselling it in competition with the ordinary retailers.

This outcome could be averted if the manufacturer were permitted to contract around the first sale doctrine. A product such as an iPad comprises numerous patented components and is loaded with copyrighted software. If the first sale of the iPad does not exhaust these patents and copyrights, arbitrage could be prevented more effectively because every unauthorized reseller would infringe those rights. So far so good. However, if the sale of the iPad does not exhaust these patents and copyrights, any further transfer of the iPad would require the permission of the relevant patent and copyright owners (which may or may not be the producer) for the entire duration of the patents or the copyrights. Consequently, the buyer would not be permitted to sell her old iPad a few (or many) years down the road, not even when the product becomes obsolete. It is evident that such long-lasting restrictions, which apply \emph{in rem}, are not necessary for solving the original problem, namely opportunism among retailers. The restrictions will remain even when opportunism is no longer a concern. At the same time, the duration of such restrictions and their \emph{in rem} applicability may impose diverse costs on future buyers and users (among others), which the
producer will not internalize.

Therefore, while the availability of remedies for IP infringement reduces the costs of enforcing post-sale restrictions, such remedies may increase the social costs associated with such restraints (more on which below). This leads us to the next proposition:

**Proposition 3:** The economics of joint production do not normally justify post-sale restrictions imposed on third parties. Therefore enforcing post-sale restraints should not normally be part of the property bundle.

Proposition 3, of course, is crucial to the design of an optimal exhaustion rule, because if true, it would reject the weaker forms of exhaustion. Therefore, it is important to elaborate not only on why and when post-sale restraints are efficient but also what the optimal way to enforce them is. In other words, choosing the optimal exhaustion rule requires an understanding of not only when post-sale restrictions are efficient, but also an ability to tell whether they should be enforced as a matter of contract law (and perhaps tort law), IP law, and against whom.

**A. The Costs and Benefits of Infringement Remedies**

In his recent analysis of the first-sale doctrine, Herbert Hovenkamp embraced the superiority of enforcement through infringement actions over suits for breach of contract and criticized the US government, which argued, in its amicus brief in *Quanta*, that post-sale restrictions should be enforced solely as a matter of contract law. Hovenkamp criticized the government (as well as the Court for deciding the case in a manner consistent with the government’s position) for being more concerned about form than substance,\(^{92}\) and as “excessively draconian”\(^{93}\) for yielding a *per se* rule that denies automatically any IP remedy once a sale is found, “with no consideration of the restraint’s purpose or effect.”\(^{94}\) Hoenkamp’s preference for infringement remedies is based mainly on their applicability to a large and diverse number of downstream people,\(^{95}\) on the fact that, at least historically, infringement actions incorporated a preference for injunctive relief,\(^{96}\) on the calculation of damages on the basis of a

\(^{92}\) Hovenkamp, *supra* note 15 at 51.

\(^{93}\) *Id*, at 56.

\(^{94}\) *Id*, at 51.

\(^{95}\) *Id*, at 51.

\(^{96}\) *Id*, at 53.
reasonable royalty together with interest and costs (and up to treble damages for willful patent infringement), rather than expectation damages, and on a possible jurisdictional advantage of suing in a federal rather than a state court. In addition to these legal advantages, Hovenkamp also notes that although theoretically a seller might be able to bind all downstream purchasers contractually, by demanding that the initial purchaser not only to agree to the restriction, but also, upon pain of damages, to impose the restriction on all parties downstream, this solution would be much messier and will increase transaction costs significantly. Like a servitude on real property that provides a neater and cheaper solution, enforcing a post-sale restriction as a matter of IP law would be more efficient, at least as long as, and only against, purchaser who have timely notice of the restriction.

Hovenkamp is not the first commentator to invoke the common law’s tolerance towards land servitudes to juxtapose it against the first-sale doctrine and its hostility towards post-restrictions in order to maintain that as long proper notice can be given this differential treatment demonstrates the alleged irrationality of first-sale doctrine (as well as of the traditional hostility towards servitudes on chattels). Nevertheless, the analogy to land servitudes actually helps demonstrate the inferiority of enforcing post-sale restriction with IP remedies. The economics of efficient land use and the economics of efficient distribution (and use) of goods embedding IP differ substantially. Although both land use restraints and post-sale restraints may solve organizational problems and foster more efficient asset use, the organizational problems are different and the negative effects of enforceable restraints are different. Consequently, the costs and benefits of various instruments for enforcing the restraints (e.g., contract vs. property rules) are not the same, and it would be an unfortunate error to apply uncritically lessons from one area to another.

Despite early skepticism, the common law has grown to accept the

---

97 Id, at 53-54.
98 Id, at 54.
99 Id, at 52.
100 Id, at 52-53, 56.
101 See e.g., Glen O. Robinson, *Personal Property Servitudes*, 71 U. CHI. L. R. 1449; see also Molly Shaffer Van Houweling, *The New Servitudes*, 96 GEORGETOWN L. J. 885 for a critical survey of this literature.
102 See Leah Theriault, SJD Thesis, Faculty of Law, University of Toronto (2010).
utility of land servitudes as means of private ordering to coordinate productive land use and reduce potential conflicts in ways that supplement and sometimes substitute for land-planning law.\textsuperscript{103} Increased urban density\textsuperscript{104} and the inevitable externalities that the use of one piece of land creates for others, coupled with the finite nature of land, render indispensable the need for restrictions on use. Moreover, these attributes of land suggest that generally the need for coordination is as durable as the land itself, is unlikely to decrease over time, and will often outlive the tenure of those who own or possess the land at the time the need to coordinate arises. Under such circumstances, the cost-inefficacy of reliance on contracts to enforce land-use restriction is glaring and the superiority of servitudes is striking.

At the same time, the very nature of land and the coordination and externalities problems it presents suggest that the benefits of servitudes often outweigh their costs. The common law’s reluctance towards servitudes has traditionally focused on three types of concerns: the problem of notice and information costs (that may arise when purchasers of burdened land may not be aware of restrictions limiting the use of the purchased asset),\textsuperscript{105} the problem of the future (arising from limitations on the freedom of future generations to manage resources wisely, autonomously, and efficiently),\textsuperscript{106} and concerns over other externalities (such as restraints on trade and competition or the limitation on individuals’ fundamental rights).\textsuperscript{107} The development of registration systems (together with notice-encouraging doctrines)\textsuperscript{108} and the fact that the value involved in land transactions typically justify incurring the costs of learning about potential land restrictions, have minimized the notice problem. The problem of the future may still be of concern in land, but it is precisely the need to coordinate land use over long periods of time that create the need for enforceable use-restrictions and highlight the superiority of enforcing them as servitudes rather than by contract. Therefore, it makes sense to prefer solutions for known and pressing needs arising today and affecting

\textsuperscript{103} Van Houweling, id at 892.

\textsuperscript{104} Id.

\textsuperscript{105} Id, at 893.

\textsuperscript{106} Id, at 900.

\textsuperscript{107} Id, at 905.

the foreseeable future and to discount the possibility that such solutions would create possible but undeterminable problems in the distant future. Similarly, the need to impose restraints for internalizing external costs suggests that the externalities calculus weighs in favor of land servitudes for several reasons: first, the local nature of land makes it less likely that land servitudes will systematically result in a significant anti-competitive harm, or otherwise violate important public policies. Second, when they do, antitrust law, the common law of restraints on trade, and the ability of courts to invalidate such servitudes may effectively weed-out the harmful restraints.

The cost-benefit calculus yields different outcomes when post-sale restraints of goods embedded with IP rights are concerned. From an organizational perspective, both land servitudes and post-sale restrictions may promote productive asset use in situations when integration is incomplete or absent. If one firm owned all tracts of land so that it internalized all costs, it would not need to create servitudes, because it could decide for itself the optimal way to use it. Similarly, a firm that is fully integrated into distribution and retail would not need to impose post-sale restrictions, because it could control all externalities between its units through managerial orders. Beyond these similarities, however, land servitudes and IP servitudes (post-sale restraints enforceable as a matter of IP law) are mirror images of each other.

The organizational problems that post-sale restraints can solve occur primarily within the relationships between co-producers, that is, firms who jointly collaborate to produce and efficiently distribute information goods and need to be able to control opportunistic behavior. Two important observations follow: first, the problems arise at the early stages of the product life-cycle. Production problems disappear immediately after the good is produced, and distribution problems largely disappear upon the distribution of the good or shortly thereafter. Importantly, all of these problems cease may to exist long before the IP rights expire (particularly in the case of copyright). Second, as co-producers, the firms are in privity that

---


110 Van Houweling, supra note 92 at 905.


112 Id.
enables them to rely on contracts for addressing many of the organizational problems associated with efficient distribution. Therefore, unlike land, whose efficient use may require long-term restrictions enforceable against many future buyers, the organizational problems arising in producing and distributing information goods that justify post-sale restraints are short-term and occur primarily between firms in privity, which can rely, to a large extent, on contracts to ameliorate these problems. It follows that while IP remedies may increase the enforceability of post-sale restraints, the marginal benefit of IP remedies must be weighed against the fact that they might be enforced even when they no longer serve their original purpose, and when embedded in rem, they will prevent the efficient use of the good long into the future. Another consideration is that because the standard remedy for breach of contract is only damages, and the plaintiff needs to prove actual damage, a party bound by a contractually valid post-sale restraint, may still be able to put the good to better use when no damage can be shown, or otherwise when the breach is efficient.

In sum, even though post-sale restrictions might be efficient and justifiably enforceable, it does not follow that enforcing them should be part of the bundle of property rights available to IP owners. Therefore, they should generally be imposed and enforced as a matter of contract law, not of intellectual property.

VII. THE (NON)-COASIAN FLAW

The previous Part explained why it would make sense to reject both the strongest version of exhaustion, as well as its weakest version. The strongest version of exhaustion should be rejected because it may invalidate any attempt to impose post-sale restraints, even when they are efficient. The weakest form of exhaustion ought to be rejected because it may allow enforcing post-sale restraints even when their social costs exceed their benefits. I suggested that when efficient, contracts should generally be relied on to enforce them. I emphasize “generally” for two reasons. One is that the relative advantages and disadvantages of IP remedies vs. contractual remedies is also a function of the available defenses. For example, if fair use, or other defenses (such as misuse), can do a good job in ensuring that only benign restraints are enforced, the calculus may change.

---

113 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. Id, at 141-42.
The second reason is that it is conceivable, at least in theory, that there might be situations in which an IP remedy might be superior after all. For example, if it could be demonstrated that a life-saving drug would not have been produced, or would not be sold in developing countries, unless all opportunities of arbitrage are eliminated, and that this requires IP remedies, then in theory, such remedies should be available.

Therefore, the important question is not so much about a categorical choice between IP and contract (and whether clever drafting can guarantee that a restriction is found to be a license condition rather than contractual covenant)\textsuperscript{114} but about how strong a default exhaustion rules should be. In other words, the important question is how sticky defaults exhaustion rules should be, under what conditions courts will enforce deviations from such defaults, and what the grounds for invalidating such deviations are. Equally important is the question of who bears the burden of upholding or invalidating the restraint: should the restraint be considered presumptively valid unless proven otherwise, or should contracting around the first-sale doctrine be regarded as presumptively invalid, unless the IP owners demonstrates its efficiency. In this Part I advocate the latter option, and offer the following proposition:

**Proposition 4**: Contracting around the first-sale doctrine should be presumptively invalid. Courts should refuse to enforce license conditions or contract terms limiting the ability of the user to resell goods embodying IP rights, unless the IP owner can demonstrate that the restraint is necessary and superior to other means to achieve efficiency.

To simplify the discussion I will use a framework of a default property rule, a contract that modifies it, and the question of whether the contract is valid or not. But the analysis can also be useful for identifying whether restraints should be enforced as a matter of IP but a proper defense may apply. To support Proposition 4, I need first to reject the following view, namely that exhaustion is simply a default rule, but one that can be easily modified by contracts, which would generally be held valid. In fact, current law probably reflects this view and most of the debate centers around relatively marginal questions such as what constitutes a valid contract,\textsuperscript{115}

\textsuperscript{114} MDY Industries v. Blizzard Entertainment, 97 U.S.P.Q.2D (BNA) 1001, (9th Cir 2010).

\textsuperscript{115} See e.g., ProCD v. Zeidenberg, 86 F. 3d 1447 (7th Cir. 1996) (discussing the validity of shrinkwrap licenses and holding that they are valid, as long as the buyer can return the product after having an opportunity to read the terms).
whether a first conditional sale preempts exhaustion, \(^{116}\) whether there was a sale or just a license, \(^{117}\) or whether restraints imposed by notice are sufficient. \(^{118}\)

*ProCD v. Zeidenberg* \(^{119}\) represents the classic articulation of this view. While *ProCD* is not a first-sale case, it is relevant because it stands for the proposition that IP rights create only default entitlements, which transacting parties can freely deviate from by contract, and that such contract terms are generally enforceable “unless their terms are objectionable on grounds applicable to contracts in general (for example, if they violate a rule of positive law, or if they are unconscionable).” \(^{120}\) In this decision, Judge Easterbrook was clear that deviating from the set of entitlements set by copyright law does not “violate a rule of positive law” because the public policy embedded into copyright law’s choice of entitlement cannot be affected by contracts: property is property and contracts are contracts. Unlike a copyright, which is “a right against the world,” \(^{121}\) “[c]ontracts generally affect only their parties, [and] do not create ‘exclusive rights.’” \(^{122}\)

Lastly, according to Judge Easterbrook, license terms accompanying works or products are conceptually identical to any other feature of the products. Both determine what the consumer can do with the product, and both are reflected in the product’s price and mediated through the market. Generally, courts do not design products, do not determine their prices, and therefore should not intervene in the conditions accompanying their sale. “Competition, not judicial oversight, is the best protector of consumers’ interests.” \(^{123}\)

*ProCD* applies a classic Coasian logic. By “Coasian logic” I refer to the Coase Theorem according to which in the absence of transaction costs, bargaining will lead to an efficient outcome regardless of the initial

116 *Quanta*, 553 U.S. 617 (leaving the question open).


119 *ProCD v. Zeidenberg*, 86 F. 3d 1447 (7th Cir. 1996).

120 *Id.*, at 1449.

121 *Id.*, at 1454.

122 *Id.*

123 *Id.* at 1453.
allocation of property rights. Assuming away external effects ("generally affect only their parties"), the distinction between property and contract provides a seemingly elegant answer to the question of whether there is any problem in contracting around limitations to IP rights. Assuming that such restrictions affect only their parties, the fact of a contract suggests that—save for some extraordinary circumstances—the transaction must be efficient. Assuming further that markets are perfectly competitive and consumers are fully informed, license terms can indeed be treated like any other product feature. Judge Easterbrook made a more explicit contemporaneous reference to the Coase Theorem when he stated that given the difficulty of setting optimal IP rules, the most sensible approach is to encourage Coasean bargaining by creating clear and enforceable property rights, facilitate bargaining and “enjoy the benefits.”

Applied to the first-sale doctrine, this view would uphold any contractual post-sale restriction. In a Coasean world, it does not really matter whether the first sale doctrine exists or not, because transacting parties will always be able to efficiently bargain about resale rights. It does not really matter—at least when transactions are possible—whether IP owners have an exclusive distribution right and whether it exhausts upon the first authorized sale. 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. Under the ProCD logic, we should not worry about the question of exhaustion. The law should permit and uphold all bargaining and let us all “enjoy the benefits.”

While attractive, this policy prescription is flawed. It is flawed because the world of IP is notoriously non-Coasian. 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, “The 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.” Unfortunately, it has become the world

of modern intellectual property thinking, but such logic is inherently inconsistent with the notion of intellectual property. In a Coasian world, there would be no need for either intellectual property rights or for statutorily defined limitations to such rights. The need for intellectual property rights, and the need to define their limits arises precisely because the world of information is notoriously non-Coasian.  

To begin with, in a Coasian world, there would be no need for IP rights, because prospective inventors or authors could secure the necessary return on their investment by contracting with potential users. They would pitch their ideas to potential users/investors, and those interested in using the good (if created) would undertake to finance it. Unfortunately, relying solely on contract may lead to sub-optimal level of investment because of Arrow’s information paradox: the investor may not be willing to commit to invest until after the information has been disclosed, but will have no reason to invest once the information is disclosed and he can use it for free.  

We can immediately think about other transaction costs, such as the need to transact with numerous potential investors, the fact that finance may be needed even before it is even clear what, if anything, will ultimately be invented or created and how valuable it will be, or the fact that investors may realize that they could be better off waiting until others undertake to finance, wait until the good is created and then copy it for free. Intellectual property rights are needed precisely for addressing these and similar market failures. It is therefore odd to assume zero transaction costs Coasian bargaining with respect to IP rights when the existence of such rights is based on the opposite assumption.

It could be argued that even if prior to creating the intellectual good the world is non-Coasian, it does not necessarily mean that once created and protected by IP rights, Coasian bargaining over those rights is equally impossible. But such a view would need to explain why the law defines not only IP rights, but also their limited time and scope. If post-creation the world were Coasian, there would be no need to limit the scope of IP

127 Indeed, this is true for any type of property, not only IP, see Thomas W. Merrill and Henry E. Smith, Making Coasean Property More Coasean, J. L. & ECON. (forthcoming 2011), Available at SSRN: http://ssrn.com/abstract=1758846 (“We have property and endow it with a basic architecture of exclusion rules supplemented by rules and standards governing proper use, precisely because of transaction costs.”)

The term of exclusivity will be mediated through the market and the price system. There would be no need for fair use or other limitations on owners’ rights, because the socially optimal outcome would be achieved even if owners had total control; each and any socially efficient use of the work would be authorized (“You want to criticize my work? Go ahead, here’s a license”; “You want to build on my ideas to develop a better product that will displace my own? No problem, here’s your license”). In a perfectly Coasian world there would be no problem treating what IP law allocates to owners and users respectively only as starting point from which transacting parties will freely negotiate, conclude contracts and necessarily make all of us better off. Indeed, such allocation could be totally arbitrary as long as the law clearly defines what is allocated to whom and then enforces all subsequent contracts. But if we realize that our world is not perfectly Coasian and that limitations on IP rights play an important role in such non-Coasian world, then we cannot simultaneously assume that IP law provides only the baseline from which bargaining will necessarily, or even presumptively, increase social welfare. Of course, it does not mean that contracting-out of limitation on IP rights cannot increase social welfare, because we have no reason to assume that the initial allocation is always optimal, and we have grounds to believe that sometimes it is not. But if the law creates those limitations under the assumption that transacting parties’ private costs and benefits are not fully aligned with social costs and benefits, then it logically follows that there can be no presumption that contracting around such limitation increases social welfare.

In addition, while the distinction between contract and property, in personam and in rem obligations, is a valid one, it loses force in the context of mass-marketed copyrighted works. When massively distributed works are sold with restrictive terms, and it becomes difficult to access works to which such terms do not apply, the practical result may be quite similar to an in rem obligation. When technological measures reinforce the restrictions and the law prohibits circumventing them, the distinction collapses altogether, because the restrictions affect everyone in touch with the work—even those who are not privy to the initial contract or the license.

129 Compare Merrill & Smith, supra note 127, at ___ (“In a world of zero transaction costs, it would not matter whether property rights are broad or narrow, clear or ambiguous – or in rem or in personam.”)

Moreover, Judge Easterbrook’s characterization of license term as product features generally immune to judicial oversight does not really support the weight of the laissez faire conclusion he arrives at. While it is true that as a general matter competition protects consumers’ interests better than judicial oversight, there are at least three reasons why market competition may not create conditions that generally justify a presumption of efficiency for contracting out of IP limitations.

Firstly, by design, markets for copyrighted and patented works are not perfectly competitive. Only the owner can legally be the source of copies of any particular work. There cannot be unfettered competition between different sellers of the work, competing over price and license terms. Although competition from other non-infringing works may still exist, assuming that this competition resembles the conditions in markets with near-perfect competition is conceptually flawed and empirically incorrect. If that were true, IP rights would be totally useless.  

Second, even if license terms can be conceived as equivalent to product features, the more complex the combination of features/terms/prices is, the more likely it is that consumers will not be fully capable of fully apprehending what they are contracting on. If such information asymmetries exist, the assumption that the market functions efficiently becomes less credible. While it is true that competitors might be interested in bridging the information gap in order to increase their own sales at the expense of their rivals, competitors may not always find it worth their while to bridge the information gap, and as noted above, IP law intentionally reduces the intensity of competition anyway.

Thirdly, even when competition exists and users are fully informed about the rights they waive, such transaction may not always be efficient because the transacting parties will fail to consider the externalities imposed on third parties. Many users’ rights (including copyright subject matter limitations) permit users to engage in creative activities that benefit not only


133 It should be mentioned that trademark law may also reduce competitors’ ability to inform their consumers about the advantages of their products or services over those of their rivals, see Ariel Katz, Beyond Search Costs: The Linguistic and Trust Functions of Trademarks 2010 BYU L. REV. 1555 (2010).
them, but third parties and society at large.\textsuperscript{134} These users, however, can expect to internalize only part of the social benefit arising from their use; they do not take into account the positive spillovers conferred on others.\textsuperscript{135} Owners ignore those positive spillovers too, and would be especially reluctant to permit uses that might harm their own interests. Consequently, under conditions of full alienability of user rights, the market will fail to generate socially optimal transactions between owners and users.

Lastly, even if follow-on innovator and earlier innovators could internalize all such spillovers, they may still not be able to negotiate effectively because of the same information gaps that prevent efficient contracting between investors and the earlier innovators. If we cannot assume that conditions for Coasian bargaining over first and follow-on innovation exist, there is no reason to assume that contracts restricting such innovations are efficient.

The previous Part challenged the wisdom of making IP remedies generally available for enforcing otherwise valid restraints, noting that IP owners and their transacting parties may not internalize the social costs arising from the long term and \textit{in rem} nature of IP entitlements. It suggested that enforcing post-sale restraints should generally be within the domain of contracts. This Part went further and warned against adopting a false Coasian view, namely that contracting-around IP limitations is presumptively efficient socially. The following Part ties together these two insights specifically in the context of exhaustion, and offers a justification for the continuing vitality of the first sale doctrine.

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{134} Feist Publications, Inc. v. Rural Telephone Service Co., Inc., 499 U.S. 340 (1991); CCH Canadian Ltd. v. Law Society of Upper Canada, 2004 SCC 13, at para. \textit{Feist}; CCH, at para. 23 (recognizing “society’s interest in maintaining a robust public domain that could help foster future creative innovation”). \textit{See also} Théberge v. Galerie d’Art du Petit Champlain inc., 2002 SCC 34, [2002] 2 S.C.R. 336, at para. 32 (“Excessive control by holders of copyrights and other forms of intellectual property may unduly limit the ability of the public domain to incorporate and embellish creative innovation in the long-term interests of society as a whole, or create practical obstacles to proper utilization. This is reflected in the exceptions to copyright infringement enumerated in ss. 29 to 32.2, which seek to protect the public domain in traditional ways such as fair dealing for the purpose of criticism or review and to add new protections to reflect new technology, such as limited computer program reproduction and “ephemeral recordings” in connection with live performances.”)

\end{itemize}
\end{footnotesize}
VIII. Justifying the First Sale Doctrine

Recall that the need for (socially efficient) post-sale restraints arises primarily in settings of joint-production and incomplete vertical integration. In such cases working around exhaustion rules is not objectionable, at least to the extent that contracting around guarantees or improves production. In contrast, restraints in settings that do not involve joint-production (e.g., those imposed on end users and affecting third parties) might result in a different cost-benefit calculus. While IP remedies for enforcing such restraints might be more effective, the social costs of IP remedies might outweigh the benefits. Also recall that the costs and externalities endemic in contracting around information suggest that there is no reason to presume that contracting around IP limitations will be socially efficient. In this Part I focus more closely on justifying the first sale doctrine. I show that the first-sale doctrine improves welfare by reducing transaction costs and explain why even when full notice is given, IP owners and their transacting parties will internalize neither the full costs of their transactions not the full benefits that exhaustion permits.

There are at least two types of social costs that IP owners and their transacting parties would likely ignore when they want to maximize the private benefits of their transactions: static and dynamic. The main static costs relate to preservation and resource waste of the physical objects, while the main dynamic costs involve loss of knowledge and negative impact on future innovation, particularly user-innovation. 136

A. Preservation and Resource Waste

Copyright owners discontinue a large number of books and recordings each year. Of the more than ten thousand books published in the United States in 1930, only 174 were still in print as of 2002. 137 And in 1999 alone, Barnes and Noble stated that 90,000 books went out of print. Many of these books are shelved in public libraries and private domiciles. Few remain in publishers’ warehouses. 138

136 Theriault, supra note 102.


Losses in music and film are equally dramatic. Only about twenty percent of feature films from the 1920s and fifty percent of feature films made prior to 1950 still survive, whereas sixty percent of all sound recordings are now estimated to be out of print.\(^{139}\)

The first sale doctrine assures that artifacts embedding works remain available to the public over time.\(^{140}\) It achieves this “availability effect”\(^{141}\) in at least three ways. First, the exhaustion doctrine makes it legal for third parties to supply copies of a work when a copyright owner ceases to make the work available. This establishes a secondary market that not only provides access to intellectual goods at lower prices, but also increases the chances that copies will be disseminated to a wider variety of people and locations beyond the original and often narrow distribution streams.\(^{142}\)

Second, exhaustion offers an incentive to purchase copies in the first place. Consumers may be more willing to purchase an item if they know they can later resell the copy in a secondary market.\(^{143}\) In addition to being consistent with the classic incentive paradigm for IP, this helps ensure preservation by reducing the opportunity cost of acquiring an intellectual works that one might otherwise forego, such as an obscure book or record that is released with little marketing or fanfare.

Finally, the first sale doctrine contributes directly to the survival of copies of works or patented goods over time by discouraging abandonment and waste. The ability to sell or donate a used copy of an intellectual good to another person inhibits the natural tendency (or the legal need) to discard an item when keeping or preserving it is costly but the item is no longer useful, convenient, or economic to keep.\(^{144}\) In short, the first sale doctrine

---


\(^{141}\) Reese, supra note 121 at 592.

\(^{142}\) Id. at 607

\(^{143}\) Id. 607

\(^{144}\) Id. at 607-608
enshrines preference for the garage sale over the garbage bin, for the library over the landfill.

In addition to these static benefits, the first sale doctrine contributes to dynamic efficiency by permitting secondary market channels that enable works and the ideas they carry, or goods and the technologies embedded therein, to remain accessible to the public even if the copyright holder ceases production or distribution of the work.\(^{145}\) It creates a built-in back-up system that ensures public access to history and culture notwithstanding the private bargain between the IP holder and her immediate licensee.

**B. User Innovation**

Implicit sometimes in the debate about exhaustion is the argument that whatever benefits IP owners ultimately benefits society because these benefits provide greater incentives to create, and sometimes, greater benefits to disseminate. An extreme version of this argument would be that even if post-sales are imposed for the sole reason of increasing the IP owner’s market power, there is nothing objectionable about that because the additional profit will benefit society in the longer run.\(^{146}\) Such view is consistent with the predominantly producer-centric model of innovation. In this model, the most important designs for innovations originate from producers and are supplied to consumers via goods and services for sale. The higher the profit available, the greater is the incentive to innovate. However, recent research on user innovation challenges this nostrum. It shows that innovation often occurs outside the firm. This suggests that sensible innovation policies may require more than just ex ante justification; they must also focus on conditions that emerge ex post.

The producer-centric model of innovation, which predominates current debates around IP and antitrust, confronts a moving empirical reality. As Eric von Hippel observes, “innovation by individual users and open collaborative innovation are modes of innovating that increasingly compete with and may displace producer innovation in many parts of the

\(^{145}\) Hinkes, *supra* note 120 at 689.

\(^{146}\) *Cf.* Verizon Communications Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398 (2004) (“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.”, *id*, at 407).
Thus, there is not one model of innovation. There are three: one driven by producers, one facilitated by users, and one created through open collaboration.

User innovation is driven by lead users having an intention to use rather than sell an innovative technology. It arises where design and transaction costs are low. The distinguishing characteristic of this model is that lead users typically reveal their designs for free, without profit motive, at least not an immediate one. This makes user innovation more desirable over alternative methods of innovation when it is technologically feasible; its openness imposes less deadweight loss on consumers, while minimizing transaction costs for other future producers.

Many user innovators are lead users. These users have heterogeneous needs not easily satisfied by mass-market production. Their unique intellectual capital and idiosyncratic needs and experiences often allow them to encounter problems and devise solutions that no one else can. They frequently anticipate features that improve products but for which no general demand yet exists. Lead users are thus distinguishable from the general consumer in that they possess an expertise and need for optimization that impels modification of an existing product for their own use. The common feature among user innovations is that they emerge from many small ideas, obtained through use and manipulation of an asset, over long periods of time.


149 Baldwin and von Hippel, supra note 127 at 27.

150 ERIC VON HIPPEL, DEMOCRATIZING INNOVATION 34-43 (2005).

151 Leah Marie Theriault, A User Innovation Theory of the Numerus Clausus (2010) (unpublished S.J.D. dissertation, University of Toronto Faculty of Law) (on file with author) at 51-66; see also SCOTT PAGE, THE DIFFERENCE 31, 153-154 (Princeton 2007) (on the relevance of the individual’s perspective, heuristics and toolbox to solving innovative problems); VON HIPPEL, supra note 131 at 147-64 (on users having different ‘toolkits,” which allow them to match their skill to a particular problem).

152 Theriault, supra note 132 at 57-58.

153 R. Keith Sawyer, Creativity, Innovation and Obviousness, 12 LEWIS & CLARK L.
If non-producer centric models of innovation are becoming important sources of innovation, at least as important as producer-centric models, then the design of optimal exhaustion rules should take account of how they affect the various sources of innovation. Exhaustion fosters user innovation because it allows a user who possesses a copy or an artifact to experiment with it without the need to obtain permission from the IP owners, and it further fosters user innovation because it also permits users who do not innovate to resell or otherwise transfer the goods to users who do innovate.154

The Coasian view described in the previous Part is influenced in part by the notion that inventors must be able to capture the full social benefit of their innovation for there to be adequate incentive to innovate. Moreover, unless and until an inventor fully internalizes the cost of innovation, he or she will not be able to efficiently manage an innovation after it is created.155 The presumption is that the inventor is the most informed about the value of an invention, and that value can be captured fully by the transaction price.

User innovation challenges this thesis because it shows that few buyers and sellers have any idea at the time of invention or at the point of sale what innovations will ultimately emerge. If innovation is tied to asset use, it is impossible for the seller to account fully for the asset’s value at the time of sale. Few can predict what innovation will follow. Fewer still can predict what value such innovation may have for social, cultural, and political life.156 The best the seller can do is to price the asset according to a cost-benefit analysis at a static point in time, a time where the probabilities of benefits are innately unknowable but the costs are immediately felt.

User innovation also shows that who is best able to maximize the value of an asset also remains unknowable at the time of transaction. It suggests

---


155 This view is consistent with the work of Harold Demsetz on externalities. See Harold Demsetz, Toward a Theory of Property Rights, 57 AM. ECON. REV. (PAPERS & PROCS.) 347 (1967); WILLIAM J. BAUMOL, THE FREE MARKET INNOVATION MACHINE 5 (2002). For a critique of this argument, see Frischmann and Lemley, supra note 117 at 266.

that the best-informed party, the most capable steward of innovation, may not be who we traditionally think and in fact may not be a single entity at all. Therefore, in a truly Coasian way, if we cannot tell ex ante who is in the best position to further innovate, rules allowing possessors to innovate reduce transaction costs.\footnote{A proof for this conclusion may be found in the network neutrality literature, for example, where innovation is said to proliferate precisely because the Internet is a general purpose technology whose end-to-end architecture eliminates transaction costs among user innovators. \textit{See} Barbara Van Schewick, \textit{Internet Architecture and Innovation} 67-111, 357-361 (2010) (analyzing the end-to-end argument’s relationship to innovation); Jonathan Zittrain, \textit{The Future of the Internet and How to Stop It} 70 (2008) (on the importance of “generative” systems for innovation); Richard G. Lipsey, Kenneth I. Carlaw, and Clifford Bekar, \textit{Economic Transformations: General Purpose Technologies and Long-Term Economic Growth} 1-219 (2005) (on growth, technological change, and general purpose technologies).}

Restrictions on post-sale use are attempts by producers to internalize or capture as much of the social benefit from an innovation as is possible. They are an attempt to limit spillovers through contract. Contrary to full-appropriation thesis, there is evidence that the spillovers from innovation actually encourage more innovation.\footnote{Frischmann and Lemley, \textit{supra} note 117 at 278.} Accordingly, the most efficient means of fostering innovation may be to allow knowledge spillovers by inhibiting post-sale restraints. As Lemley and Frischmann explain, “‘If a technology might be repurposed or improved in ways that the initial innovator is unlikely to foresee, full internalization will interfere with the socially optimal development and use of a technology.’”\footnote{Id.}

In sum, the first-sale doctrine improves welfare by facilitating efficient use of physical and intellectual assets (including human capital) both statically and dynamically. However, IP owners and their immediate transacting parties—seeking to maximize their short-term private benefits—are unlikely to fully internalize the social costs of their transactions and therefore may impose or agree to restraints that are less than socially optimal, even when full notice is given. Therefore, contracting around the first-sale doctrine should be met with a healthy dose of legal skepticism and the burden of demonstrating the efficiency of the restraints should fall on the transacting parties. When the restraints purport to bind third parties and have long-term effect, the dose of suspicion should increase.
IX. APPLICATION

This Part demonstrates how the framework developed in this Article bears on recent cases involving the first-sale doctrine.

A. Costco v. Omega

Costco v. Omega was a parallel trade case, and seems to conform the paradigmatic international price-discrimination story. Costco purchased from a third party importer Omega watches that were originally sold by Omega to distributors in Egypt and Paraguay. Costco sold the watch for $1,300 instead of the $2,000 suggested retail price. Omega sued Costco for copyright infringement. The copyright in question was in a logo engraved on the back on the watch. The Ninth Circuit agreed with Omega, holding that the first sale doctrine did not apply because although under 17 U.S.C. § 109(a), the owner of any particular copy "lawfully made under this title" may resell that good without the authority of the copyright holder, “lawfully made under this title” means lawfully made in the United States. Because the copies were made outside the United States the US distribution right had not been exhausted by the authorized foreign sales and the unauthorized importation constituted infringement of the copyright.

Whether or not the Ninth Circuit correctly interpreted § 109(a) exceeds the scope of this paper. But from an economic point of view, the holding makes little sense. Obviously, Omega was interested in preventing arbitrage that interfered with its international distribution system. It does not follow, however, that IP law should be relied on to prevent this type of arbitrage. As discussed in Part V.C, even if price discrimination is desirable, there is little economic justification to treat domestic price discrimination differently from international price discrimination. Therefore, the only sensible choice is between universal exhaustion and no-exhaustion at all.\(^\text{160}\)

Moreover, a legal rule that treats the circulation of goods differently depending on the question of where a copy of work was made (let alone the logo) may create perverse incentives. Instead of deciding where to produce copies on the basis of economic factors, firms may shift production of copies abroad in order to gain greater control over the downstream distribution of their goods. Again, if such greater control is desirable it should be granted regardless of the place where the copy was made. However, when considering whether such control is desirable or not, it is

\(^{160}\)Supra, Part __
important to consider the relative merits of different legal tools to exert it.

As Parts VI to VIII demonstrate, while allowing IP owner to invoke their IP rights to tightly control the distribution of their goods may improve the efficiency of their distribution systems, IP remedies are not generally essential for achieving such efficiency, because contractual remedies might be sufficient. It is not enough to show that relying on IP remedies is more efficient than relying on contractual remedies. To justify the availability of IP remedies it must be shown that the marginal benefit of IP remedies over contractual remedies outweighs the social cost associated with IP but lacking with contract.

B. Vernor v. Autodesk

Vernor v. Autodesk involved the resale of used copies of software. Timothy Vernor purchased several used copies of Autodesk, Inc.’s AutoCAD Release 14 software (“Release 14”) from one of Autodesk’s direct customers, and then resold them on eBay. He then brought a declaratory judgment action against Autodesk to establish that these resales did not infringe Autodesk’s copyright. Vernor argued that the first-sale doctrine permits the sale of those used copies. The 9th Circuit disagreed. The court reasoned that “[i]n its current form, [the first sale doctrine] allows the “owner of a particular copy” of a copyrighted work to sell or dispose of his copy without the copyright owner’s authorization.” The question, therefore, was whether Autodesk’s direct customer (who sold the copy to Vernor) was an “owner of a particular copy”. The Court found that the customer was not an “owner” of the copy but mere “licensee”, because when the customer installed the software it had to accept Software License Agreement in which Autodesk (1) specified that the user is granted a license, while ownership in the copy remained with Autodesk; (2) it significantly restricted the user’s ability to transfer the software; and (3) imposes notable use restrictions. Finally, the court noted “the significant

---

161 Vernor sought a declaratory judgment after on several earlier occasions Autodesk filed Digital Millennium Copyright Act (“DMCA”) take-down notices with eBay claiming that Vernor’s sale infringed its copyright, and eBay terminated Vernor’s auctions. In each of these cases the auctions were reinstated after Vernor filed counter-notices and Autodesk failed to respond. At one point, however, eBay suspended Vernor’s account for a month because of Autodesk’s repeated allegations of infringement, during which Vernor was unable to earn any income eBay, id, at ... 

162 Id, at

163 Id, at
policy considerations raised by the parties and amici on both sides of this appeal, but decided that the as a matter of statutory interpretation and precedent such considerations should be left to Congress. As a result, by requiring users to “agree” to the Software License Agreement, Autodesk was able to impose post-sale restraints on copies of its software.

To determine whether this outcome is desirable, it is useful to understand what Autodesk tried to achieve by preventing the resale of its software. One possibility is a simple price discrimination scheme: according to the court Autodesk offers the software with different terms for commercial, educational institution, and student users. Arguably, maintaining this scheme requires an ability to prevent arbitrage, e.g., preventing a student user from reselling her copy to a commercial user, and obviously, preventing arbitrage is easier when the unauthorized sale of a lawfully made copy constitutes copyright infringement.

The case, however, did not involve a student or an educational user reselling her copy to a commercial user but a commercial user upgrading his software to a newer version and reselling the old one. Most likely, Autodesk sought to prevent the resale to maintain another type of price discrimination. When Autodesk released Release 15 it sold it for $3,750 apiece, meaning that a new commercial user must pay this amount if Release 15 is the only game in town. An existing user of Release 14, however, can continue using that earlier version, even if the new version is better. In other words, existing users have more elastic demand for the new version than new users and in order to convince existing users to upgrade, Autodesk offered them the new version for a substantially lower price: only $495. Obviously, if the upgrading users can resell the older version to others, Autodesk’s new version is no longer the only game in town. Rather than paying $3,750 a new user has the option of buying a used copy of the older version, meaning that like existing users, she may only be willing to pay $495 for the new version.

As always with price discrimination, the welfare implications of such a scheme are ambiguous. Clearly, Vernor’s engagement in the resale of used

---

164 Id., at
165 Id., at.
166 Id., at
167 I assume away the availability of competing programs. This assumption is not unrealistic, as Autocad is currently the de facto standard in this type of programs, with a market share of 80% (cite).
copies indicates that the quantity sold and prices set by Autodesk do not satisfy the entire demand for Autodesk, and it is far from clear that by eliminating the secondary market Autodesk will increase output. Therefore, it is not clear whether allowing Autodesk to control the secondary market increases or decreases allocative efficiency. There is no clear answer about the effect on dynamic efficiency. Evidently, Autodesk’s attempts to eliminate Vernor strongly suggests that the secondary market eats into its profit, and this can be a bad thing if lower profits reduce the incentive to innovate. On the other hand, it is equally possible that competition from the secondary market drives Autodesk to keep innovating, rather than selling the same version for years until the market is fully saturated.

But even if one accepts that Autodesk should be allowed to prevent a user who purchases a new version at a lower price from reselling the old one, it is far from clear how far downstream it should be able to control such resales if the user breaches. In other words, the presumed efficiency of restricting the immediate user does not immediately translate into a property right against any subsequent user. Given that permitting Autodesk to design its initial transaction in a way that allows it to control any subsequent transaction can have long-term and unknown external costs that neither Autodesk nor its immediate user internalize, it would be useful to know whether a remedy for infringement is commensurate with the problem requiring solution. Even if the existence of a secondary market is problematic, the problem is short term, mainly during the transition from one version to a newer one. Moreover, Autodesk and the upgrading user can be in contractual privity. Autodesk could have sued the user for breach of contract and recover its damages if they existed.\(^\text{168}\)

\(\text{C. UMG Recordings v. Augusto}\)

The last of the recent Ninth Circuit cases involving the first sale doctrine is UMG v. Augusto. UMG is one the world’s largest music companies.\(^\text{169}\) Like many music companies, UMG ships specially produced “promotional CDs” to a large group of industry insiders, such as music critics and radio

\(^{168}\) Interestingly, the decision in Vernor indicates that Autodesk sued not only Vernor but also CTA, the customer who sold the copies to Vernor. The parties stipulated to entry of a permanent injunction against CTA from directly or contributorily infringing Autodesk’s copyrights. There is no mention of any damage award for either breach of contract or copyright infringement, suggesting that proving damages had not been easy.

\(^{169}\) UMG v. Augusto, at ..
programs. Many of the promotional CDs bore a statement such as the following: “This CD is the property of the record company and is licensed to the intended recipient for personal use only. Acceptance of this CD shall constitute an agreement to comply with the terms of the license. Resale or transfer of possession is not allowed and may be punishable under federal and state laws.”

Troy Augusto was not a recipient of such CDs but was able to obtain numerous promotional CDs, which he then sold on eBay. UMG brought action against Augusto, claiming that selling the CDs infringed its distribution right. The Ninth Circuit affirmed the district court’s grant of summary judgment in favor of Augusto. The court held that the first sale doctrine applied in these circumstances. The distribution of the promotional CDs constituted a sale, within the meaning of § 109(a) of the Copyright Act. As a result, the recipients became owners of the copies and were free to dispose of them as they saw fit. Therefore, the subsequent sale by Augusto did not infringe UMG’s copyright. In reaching this conclusion, the court distinguished Vernor, finding that although UMG placed written restrictions in the labels of the CDs, it had not established that the restrictions on the CDs create a license agreement rather than a sale.

Unlike Costco or Vernor, UMG v. Augusto does not seem to be a case of price-discrimination. The promotional CDs are sent to industry insiders before the CDs are released for purposes of promoting and advertising the release of the new CD. The promotional CDs are similar to the new CDs, although they may contain fewer songs and may not include the artwork included with the new CD.

Although I agree with the Ninth Circuit ultimate holding, Augusto is a case that gets very close to being one that may actually justify the restraints which UMG attempted to impose on the recipients of the promotional CDs. The key difference between Augusto and a case like Vernor is that the CDs are sent before the CDs are released. The different is crucial because rather than consumers, the insiders receiving the CDs can be regarded as co-producers, at least to some extent.

Playing songs remains the most effective means for promoting them and

170 Id., at ..

171 Id. The court also found that the recipients were free to dispose of them as they saw fit under the Unordered Merchandise Statute, 39 U.S.C. § 3009.

radio stations play a crucial role in creating and shaping the demand for new songs. The relationships between record companies and the insiders who receive the promotional CDs are symbiotic. The feedback received from the insiders can be important for making the last decision before the release of a new CD, and the timing and manner in which the promotional CDs are distributed provide important signals for the insiders about the promotional efforts the record company intends to exert. These signals in turn help radio stations better forecast the changing tastes of their listeners. But sending the promotional CDs involves some risks. For example, premature “leaks” may undermine the carefully planned scheduling surrounding the release of new CDs, or may they contain materials that will not or may not ultimately be included in the released disk. All of this suggests that there might be legitimate reasons for restraining the free distribution of promotional CDs, at least temporarily.

But the temporal issue is crucial. Even if contractual obligations may not be sufficient to guarantee the optimal control when it is necessary, UMG’s notice and legal theory is overkill. The notice indiscriminately prohibited any sale or transfer of possession, even transfers occurring long after the reasons justifying the restrictions no longer exist, and even by people that cannot be regarded as co-producers in any way. In contrast, over time, the first sale doctrine contributes to the preservation of such promotional CDs, which may be socially valuable in the long run.

X. 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, and somewhat counter-intuitively, insights borrowed from modern antitrust law and economics are invoked to provide a seemingly robust theoretical foundation for undermining exhaustion rules or narrowing their scope, and thereby strengthen IP owners’ control over downstream distribution and use of the goods they produce. Just as antitrust law has recognized the efficiency of post-sale restraints and relaxed its

---

173 Ariel Katz, the Potential Demise I, at ; RIAA Amicus Brief in Augusto, at
175 RIAA brief at 10.
176 Compare, Ariel Katz, Substitution and Schumpeterian, supra, at 140-47.
hostility towards them, so should IP law permit their imposition and provide remedies for their breach.

This article shows why this trend is misguided and should be resisted, not because the insights from modern antitrust are irrelevant, quite the contrary. However, as this article demonstrates, the insights from modern antitrust do not support the case against the first sale doctrine. Upon closer examination, it turns out that the main benefits of post-sales restriction involve situations of imperfect vertical integration between co-producing or collaborating firms, and occur during the production and distribution phases or shortly thereafter. In such situations, contracting around the first-sale doctrine should be permitted. 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. Thus, when taken seriously, the insights from modern antitrust law and economics ultimately support its continued vitality.