1-1-2011

An Economic Defense of Flexibility in IPR Licensing: Contracting around First Sale in Multilevel Production Settings

Anne Layne-Farrar

Follow this and additional works at: <http://digitalcommons.law.scu.edu/lawreview>

Part of the [Law Commons](#)

Recommended Citation

Anne Layne-Farrar, *An Economic Defense of Flexibility in IPR Licensing: Contracting around First Sale in Multilevel Production Settings*, 51 SANTA CLARA L. REV. 1149 (2011).

Available at: <http://digitalcommons.law.scu.edu/lawreview/vol51/iss4/4>

This Article is brought to you for free and open access by the Journals at Santa Clara Law Digital Commons. It has been accepted for inclusion in Santa Clara Law Review by an authorized administrator of Santa Clara Law Digital Commons. For more information, please contact sculawlibrarian@gmail.com.

AN ECONOMIC DEFENSE OF FLEXIBILITY IN IPR LICENSING: CONTRACTING AROUND “FIRST SALE” IN MULTILEVEL PRODUCTION SETTINGS

Anne Layne-Farrar*

I. INTRODUCTION

In the debate that swirled just before and after the release of the U.S. Supreme Court’s opinion in *Quanta Computer v. LG Electronics, Inc.*¹ during the summer of 2008, it was clear that the basic principle of first sale was not so much at issue; rather the Court sought to delineate the doctrine’s metes and bounds. While the delineation of boundaries is primarily a matter of law, and thus best left to judges or to lawyers, one specific boundary has considerable economic implications. The particular issue of whether firms should be allowed to contract around the first sale doctrine when multiple levels of production are involved—in essence, negating the doctrine’s application through the use of specific terms and conditions in a contract agreement—is as much an economic question as it is a legal question. I address this economic question in this paper.²

* Anne Layne-Farrar is a Vice President with Compass LexEcon. This paper extends ideas first developed in a theoretical paper co-authored with Gerard Llobet and Jorge Padilla. Dr. Layne-Farrar wishes to thank Eric Goldman, Gerard Llobet, and the participants of the Exhaustion and First Sale in IP Symposium at Santa Clara University for helpful comments and suggestions, and Dhiren Patki and Alina Marinova for research assistance. The ideas and opinions, as well as any errors, in this paper are exclusively her own. Comments should be sent to alayne-farrar@compasslexecon.com.

1. *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008). The court found that patent exhaustion does indeed apply to method patents, and thus the doctrine applied to LG Electronics’ patents, as embodied in Intel’s components sold to Quanta Computers. *Id.* at 631.

2. The primary focus in this paper is patents, although I consider some analogous situations relevant for copyrights and Droit de Suite, an IP doctrine

My decidedly non-legal reading of the Supreme Court's decision in *Quanta* asserts that the Court did not rule out such contractual flexibility for patents. The decision appears to leave open the question of contractual restrictions, stating only in footnote seven that "[w]e note that the authorized nature of the sale to Quanta does not necessarily limit LGE's other contract rights. LGE's complaint does not include a breach-of-contract claim, and we express no opinion on whether contract damages might be available even though exhaustion operates to eliminate patent damages."³

Similar arguments can be made with respect to copyright, where the First Sale doctrine has faced considerable controversy now that copyrighted materials can be "sold" online. While there is no multilevel production during the sale of a copyright, there can be a chain of buyers. Before the advent of the Internet, copyright holders could count on the time and expense of making a physical copy of their protected works to restrict end buyers from doing so. This barrier limited the spread of unauthorized copies and thus maintained compensation through the sale of authorized first copies. Today, however, that self-policing mechanism is no longer reliable: end consumers can and do make copies of digital material for wide distribution on the Internet. This means a copyright holder may not receive anywhere close to a market rate of return on his material.⁴ This dynamic has led to a running debate between rights holders and consumers. Rights holders want to impose technological means of restricting users' ability to copy and distribute protected works online.⁵ Consumers, on the other hand, believe that the unfettered rights in place in the physical world, including first sale, should translate in full to the online world.⁶ For

applicable to the sale of fine arts in certain jurisdictions.

3. *Quanta*, 553 U.S. at 637 n.7 (citing *Keeler v. Standard Folding Bed Co.*, 157 U.S. 659, 666 (1895)).

4. For a discussion of the impact that digital technology has had on copyright, see generally CONGRESSIONAL BUDGET OFFICE, COPYRIGHT ISSUES IN DIGITAL MEDIA (2004) available at <http://www.cbo.gov/ftpdocs/57xx/doc5738/08-09-Copyright.pdf>.

5. It is widely accepted for online publishers and device manufacturers to use a variety of access control and tracking techniques (known as DRM) to restrict the access and distribution of digital content. See ELECTRONIC FRONTIER FOUNDATION, *Digital Rights Management*, <http://www.eff.org/issues/drm> (last visited Feb. 14, 2011).

6. Opponents of DRM argue that the digital restrictions infringe on rights

example, e-book reader programs like Adobe Reader and Microsoft Reader do not allow the user to print, copy, or paste the online material from one online account to another online or offline repository, even though those restrictions have faced stiff opposition by purchasers.⁷

Finally, note that some jurisdictions take a flexible approach to the application of first sale for works of art, one that is analogous to relaxing the copyright first sale doctrine. Specifically, consider the European doctrine known as *Droit de Suite* (DDS). The DDS rule guarantees a royalty to the creator of a work of art based upon the price obtained in its future resale. California's equivalent of DDS, for instance, provides the artist with five percent of the resale price of a covered work, so long as that price exceeds both \$1,000 and the price initially paid by the reseller.⁸ Rather than limiting the creator to the revenues available from the first sale of the artwork, DDS allows the artist (and his or her heirs, for up to twenty years after the artists' death) to benefit financially from the resale of an original painting, sculpture, drawing, or work in glass.

In all three of these instances, advocates for a strict application of the first sale doctrine have argued that allowing rights holders to charge more than one party—either

that copyright purchasers have in the offline world. See, e.g., Richard Watt, *The Past and the Future of the Economics of Copyright*, 1 REV. ECON. RES. ON COPYRIGHT ISSUES 151, 164 (2004) ("DRM may also violate certain private rights, as for example is the case of copy-proof CD Roms, since there exists the right of private copy in almost all countries."). See also David Bach, *The Double Punch of Law and Technology: Fighting Music Piracy or Remaking Copyright in a Digital Age?*, 6 BUS. & POL., no. 2 at 1, 15–16 (2004)

Whereas it was not feasible to charge newspaper readers for a photocopied article, it is feasible to charge a reader a fee to access an article e-mailed by a friend. Similarly, whereas the recording industry could not reasonably charge for a tape recording off the radio, online services such as MusicNet and Rhapsody can fairly easily charge consumers for every track they want to burn onto a CD, and they are doing just that. The double punch thus clearly raises a critical question: do previously tolerated fair use exceptions constitute basic consumer rights, or were they merely an economic deadweight loss that new digital technologies help eliminate?

Id.

7. Typically, DRM protests come in the form of "computer hacks" that enable buyers to do that which the seller is attempting to prohibit digitally. See, e.g., *Amazon's Kindle has Copyright Protection Hacked*, BBC NEWS (Dec. 29, 2009, 8:40 AM), <http://news.bbc.co.uk/2/hi/technology/8428126.stm>.

8. CAL. CIV. CODE § 986 (Deering 2010).

along a vertical production chain for patents, or along a time horizon for copyrighted material and works of art—amounts to “double dipping.” In other words, the rights holder will charge twice for the use of the same rights, thereby extracting “too much” from rights users.⁹

The foundation for the concern over double dipping is easy to understand. Once a rights holder has sold a good embodying the covered rights and has received an economically justifiable payment for those rights in exchange, how could a second payment be warranted? Two economic issues render the answer to this question less than obvious: economically justified reward and economically efficient licensing.

The first issue, economically justified reward, concerns the compensation a rights holder can expect to earn from the value of her creation. Specifically, the license fees (or sales price) received for a protected property should be the market determined value. This value should be computed without resort to the “hold up” of any ex post irreversible investments a licensee or rights purchaser may have made.¹⁰

In relation to patents, it is well known that Congress enacted the Patent Act as a means of promoting “the Progress of Science and the useful Arts”¹¹ In other words, a patent is a temporary right to exclude others, and is meant to provide incentives for inventors to create “useful” technological contributions for society. A patent is therefore a legal arrangement with the government; it is not a natural entitlement. Likewise, copyright protection is also granted by the government and is typically fully enforceable only when the copyright holder has officially registered his rights.¹² And

9. For patents, *see, e.g.*, Brief of Minebea Co., Ltd as Amicus Curiae In Support of Petitioners at 4, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937). For copyright, *see* DIGITAL RIGHTS MANAGEMENT: TECHNOLOGICAL, ECONOMIC, LEGAL AND POLITICAL ASPECTS 682 (Eberhard Becker, Willms Buhse, Dirk Günnewig & Niels Rump eds., 2003). For DDS, *see* Victor Ginsburgh, *The Economic Consequences of Droit de Suite in the European Union*, 35 ECON. ANALYSIS & POL'Y 35, 61–71 (2005).

10. For a discussion of the conditions where patent hold up may occur, *see generally* Vincenzo Denicolò, Damien Geradin, Anne Layne-Farrar & A. Jorge Padilla, *Revisiting Injunctive Relief: Interpreting eBay in High-Tech Industries with Non-Practicing Patent Holders*, 4 J. COMPETITION L. & ECON. 571 (2008).

11. U.S. CONST. art. I, § 8, cl. 8.

12. *See Copyright in General (FAQ)*, UNITED STATES COPYRIGHT OFFICE, <http://www.copyright.gov/help/faq/faq-general.html> (last visited Feb. 14, 2011).

clearly, DDS is only applicable where a law is on the books—it does not naturally arise in the sale of works of art.

While neither copyright nor patent laws impose limits on what a rights holder can charge a user or licensee, the granted rights should enable the rights holder to earn enough of a financial return so as to provide strong incentives to create and innovate. This can be done by either investing in risky research and development, or engaging in costly or time consuming creative production. The rights, however, should not be so strong that they deprive society of the benefits of the innovation or grind all subsequent follow-on innovation to a halt.¹³ Intellectual property law, then, generally allows a rights holder to keep a protected creation for its own use, to refuse to license or sell the creation altogether, or to set the usage terms as the rights holder and user collectively see fit during negotiations.

In the context of economically justified rewards, the concern over double dipping is well illustrated by Justice Breyer's analogy in the *Quanta* case.¹⁴ Justice Breyer

Copyright is a form of protection grounded in the U.S. Constitution and granted by law for original works of authorship fixed in a tangible medium of expression. . . . [C]opyright exists from the moment the work is created. You will have to register, however, if you wish to bring a lawsuit for infringement of a U.S. work.

Id.

13. There is extensive literature regarding the static, as well as the dynamic, effects of intellectual property rights. Regarding the first, the trade-off between patent length and breadth has been discussed in papers such as Paul Klemperer, *How Broad Should the Scope of Patent Protection Be?*, 21 RAND J. ECON. 113 (1990) and Richard Gilbert & Carl Shapiro, *Optimal Patent Length and Breadth*, 21 RAND J. ECON. 106 (1990). The dynamic trade-offs are discussed in Jerry Green & Suzanne Scotchmer, *On the Division of Profit in Sequential Innovation*, 26 RAND J. ECON. 20 (1995).

14. Oral Argument at 1:20, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937) available at http://www.oyez.org/cases/2000-2009/2007/2007_06_937/argument. During oral argument Justice Breyer stated:

Imagine that I want to buy some bicycle pedals, so I go to the bicycle shop. These are fabulous pedals. The inventor has licensed somebody to make them, and he sold them to the shop, make and sell them. He sold them to the shop. I go buy the pedals. I put it in my bicycle. I start pedaling down the road.

Now, we don't want nineteen patent inspectors chasing me or all of the other companies and there are many doctrines in the law designed to stop that. One is the equitable servitudes on chattel. Another is the exhaustion of a patent. And now you talk about implied license.

I would say, why does it make that much difference? What we're

described firm X, holding a patent on a particular design for a bicycle pedal that it licensed for a royalty to firm Y to use the design and attach the resulting pedal to a bike.¹⁵ Justice Breyer then reasoned that when Y sells the bike to a retail shop, patent holder X cannot also charge the retail shop a royalty for the pedal design.¹⁶ Nor, when the retail shop sells the bike to a consumer, can firm X chase the consumer down the street to get yet another royalty payment for the pedal design from the bicycle rider.¹⁷

Implicit in Justice Breyer's example is the assumption that there exists one tradable right: the ability to manufacture and sell the pedal design. The same assumption is often made in regard to copyrights: a book containing a protected work is a tangible product that, once sold, passes out of the control of the rights holder. When just one right is offered, charging for any use coming further down in the production chain (say, the sale of a bike pedal to a bike assembler), or further down in the use chain (say, the resale of the book to a second hand book shop or the resale of a painting to a new collector), can be considered double dipping, charging multiple times for the same right.¹⁸

Intellectual property rights (IPR) can be thought of as a bundle of distinct rights, rather than as a single right.¹⁹ For

talking about here is whether after those pedals are sold to me under an agreement that the patent—you know, you have a right to sell them to me—why can't I look at this as saying that patent is exhausted, the patent on the pedals and the patent for those bicycles insofar as that patent for the bicycles says I have a patent on inserting the pedal into a bicycle.

Id.

15. *Id.*

16. *Id.*

17. *Id.*

18. "LGE was improperly attempting to obtain a double royalty." Brief for Petitioners at 10, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937).

19. This was an argument raised by MPEG LA LLC, as a friend of the court on behalf of LG Electronics. In particular, MPEG's brief stated that "[t]he right to make and sell a device may have one value, while the right to use that device may have completely different values to different parties." Brief for MPEG LA LLC as Amici Curiae In Support of Respondent at 24, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937). See also Stephen L. Carter, *Does it Matter Whether Intellectual Property is Property?*, 68 CHI.-KENT L. REV. 715, 716 (1993). See generally Henry E. Smith, *Intellectual Property as Property: Delineating Entitlements in Information*, 116 YALE L.J. 1742 (2007). Copyright material in digital form can also be parsed into separate rights, as

instance, there are separate and distinct rights to exclude others from using, making, or selling a good. These rights, in turn, can be split again, say into the right for others to sell in the Southern states versus the Northern states, or with a global portfolio of patents, the right to sell in the U.S. versus Europe, Asia, or elsewhere. This division of rights can be especially pertinent with intangible goods, such as for patent protected software or for a copyright protected book in electronic format.

Since IPR can embody a bundle of rights, the other pivotal economic concept in IPR licensing: economic efficiency is implicated. Just as the legal bargain offered in IPR balances incentives to innovate and to disclose that innovation against a limited term of exclusivity, it is also in society's best interest to design IPR rules so that rights holders have the incentives and the ability to sell or license their creation for broad implementation or use.²⁰ Society reaps the benefits of a new invention or a creative work mainly through diffusion and implementation. We therefore need to understand any impediments that might stand in the way of efficient licensing or selling that would hinder the diffusion and implementation of useful innovations.

One important distinction to keep in mind in considering the bifurcation of rights and their licensing, however, is the original rationale for the first sale doctrine: an increased certainty over the "price" of a good, arising when limits are placed on the parties that can be charged licensing fees.²¹ If an end purchaser of a good has no reasonable way of knowing whether the good comes with unseen obligations, such as licensing fees on the components that form the inputs of the

noted by the Congressional Budget Office:

Consumers would pay according to the particular "rights" that they are able to exercise over a copyrighted work in digital form. For example, DRM technology would prevent consumers who pay for only a few rights (say, to listen to a music file from a compact disc or the Internet) from exercising the additional usage rights (both to listen to and to make copies of the file) that are available to consumers who pay more. DRM technology could likewise be used to control consumers' ability to redistribute the copies that are made.

CONGRESSIONAL BUDGET OFFICE, *supra* note 4, at ix.

20. See generally Nancy Gallini & Suzanne Scotchmer, *Intellectual Property: When Is It the Best Incentive System?*, 2 INNOVATION POL'Y & ECON. 51 (2002).

21. R. Anthony Reese, *The First Sale Doctrine in the Era of Digital Networks*, 44 B. C. L. REV. 577, 584 (2003).

good, then uncertainty will hinder the exchange of goods and the dissemination of the innovations underlying those goods. The task, then, is to understand when relaxing first sale over a production chain will offer net benefits, versus when the cost of uncertainty calls for first sale's strict application.

I consider these issues in this paper. In particular, I consider the circumstances where flexibility in the application of the first sale doctrine to patent licensing for intermediate goods is economically efficient. I also consider, albeit less thoroughly, the sale of copyrighted works and the sale of works of art. On this point, I base my discussion on a technical companion paper that models the economics of patent licensing.²² By this approach, I find that under many realistic scenarios involving business-to-business patent licensing, social welfare is improved when a rights holder has the ability to split fees for use of the right across multiple parties. In more limited circumstances, splitting fees over users may make sense for copyrights and DDS scenarios as well. Indeed, the possibility for double dipping has nothing to do with the application of the first sale doctrine but depends instead on the rights holder's ability to exploit ex post investments in commercializing a protected creation. Therefore, it is more favorable to allow a flexible approach to first sale under certain scenarios—an approach that allows knowledgeable parties to contract around first sale's application when little or no cost of uncertainty is imposed and when it makes sense for the parties to do so.

This paper proceeds as follows. It begins by examining patent exhaustion in an unrealistic world where transaction costs and market frictions do not exist.²³ This theoretical environment provides a simple vehicle for understanding the key economic issues underlying the first sale doctrine as applied to multilevel production. With this benchmark in place, this paper then discusses how the application of the first sale doctrine might change when frictions, such as government regulations or restrictions on pricing, and transaction costs, such as the time and effort involved in

22. Anne Layne-Farrar, Gerard Llobet & Jorge Padilla, *An Economic Take on Patent Licensing: Understanding the Implications of the 'First Sale Patent Exhaustion' Doctrine* (Working Paper 2009), available at <http://ssrn.com/abstract=1418048>. All assertions made here are proven in the technical paper.

23. See *infra* Part II.

licensing negotiations, are present.²⁴ In the presence of these frictions, a strict application of patent exhaustion in licensing for multilevel production is problematic because it can lead to under-compensation for rights holders, and thus lead to a reduction in the number of transactions over protected goods. Next, this paper addresses the potential anticompetitive problems, including double dipping, which might arise in IPR licensing if the first sale doctrine allows for contractual circumvention in business-to-business licensing or other narrow circumstances with clear notification of licensing terms and conditions.²⁵ As noted above, double dipping is a distinct issue from first sale that depends on its own set of circumstances. This paper concludes by making the point that the ability to contract around first sale in the defined instances is important for economic efficiency reasons.²⁶

II. PATENT EXHAUSTION IN AN IDEALIZED WORLD

As a simple thought experiment to clarify ideas and provide a benchmark for comparison, consider first a world without frictions: all information is public and readily available, firms are free to set prices for the goods they sell and adjust them as costs and market conditions change, and negotiation among firms jointly maximizes the benefits of the parties involved. In such an idealized world, the way that IPR fees are structured for the production of goods would not matter, rendering the patent exhaustion doctrine wholly irrelevant, save for its application to end consumers.

To understand why IPR fee structuring would be irrelevant in this idealized world, consider a variation on Justice Breyer's example. A patent holder has a patent on a bike pedal design. Assume that bicycle production is split into two steps: an upstream producer makes bike components, including the pedals incorporating the patented technology, and then sells them to a downstream producer, who assembles the components into bikes that are ultimately sold to consumers. Suppose that, in accordance with a strict application of patent exhaustion, the patent holder initially only charges one firm a royalty; assume it is the downstream

24. See *infra* Part III.

25. See *infra* Part IV.

26. See *infra* Part V.

manufacturer who obtains a license to assemble and sell bicycles incorporating the patented pedal, while the upstream component maker obtains no license.²⁷ What would happen if the patent holder changed this scheme and instead began charging part of the royalty burden upstream, raising the upstream royalty from zero, but also lowering the downstream royalty so as to leave the aggregate royalty income for the patent holder unchanged?

In our ideal world with no transaction costs or frictions, the intermediate component price the upstream manufacturer charges the downstream firm would simply rise to accommodate the increase in costs that the upstream manufacturer faces. In other words, the upstream manufacturer would pass on the additional costs (the newly assessed royalty payment) to its customers, the downstream assemblers, exactly offsetting the cost-savings enjoyed by the downstream producer from the lower royalty rate. This pass-through element is the fundamental insight of economist Ronald Coase in his famous theorem on the reallocation of costs to achieve an economically efficient outcome.²⁸

But to reach this outcome, I assumed that the patent holder kept its aggregate royalty income (i.e., the sum of the up and downstream royalty revenues) the same. Do we need this assumption? Why wouldn't the patent holder instead take advantage of the multiple licensing points to increase the aggregate royalty rate? It turns out that doing so is not in the patent holder's best interests; profit maximizing incentives give us the same result without imposing any constraints on firm behavior.

When licensing just one party in the production chain—either the upstream component maker or the downstream assembler—the patent holder will set the royalty rate to maximize total licensing profits. Two forces are at play whenever a firm raises its price: first, the firm earns a larger profit margin on each unit sold in the downstream market, but second, as prices increase consumers tend to purchase fewer goods. So, firms must balance the per-unit increase

27. First sale would still apply for the sale of the bicycle to end consumers, who could not be charged royalty fees for the pedal design. See Oral Argument at 1:20, *supra* note 14.

28. See generally Ronald H. Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1 (1960).

against the accompanying decrease in units sold. If the first effect outweighs the second, the price increase will increase profits. If, on the other hand, the negative quantity effect outweighs the positive margin effect, profits will fall. The optimal price achieves the perfect balance, and thus maximizes profits. Just as they are in all profit maximization problems, these dual forces are at play for the upstream patent holder: the patent holder sets a rate to maximize its royalty earnings. Moving from licensing just downstream to both up and downstream firms does not alter the optimal royalty earnings equation; the shift simply means the sources of profit are divided. Thus, if a rights holder has found the optimal royalty earnings in licensing one production level, and then attempts to increase its profits by licensing another level, it will instead find that the decrease in the units sold outweighs the increase in the per-unit margin, and making the new higher “price” less profitable.

The presence of multiple production layers does complicate the dynamics, although the forces of profit maximization remain the same. First, within a vertical production chain,²⁹ an upstream firm will only sell its components if the downstream assembly firm demands them. This, in turn, is dictated by the demand for the final good in the downstream market. The upstream firm therefore faces “derived demand” linked to the ultimate downstream market.³⁰ Second, within a vertical production chain in our idealized world, any cost that an upstream link faces will be passed on to the downstream link, and ultimately to the end customer.

The firms facing derived demand upstream in the production process, combined with the cost pass-through noted above, means that downstream demand influences a patent holder’s licensing fees regardless of where in the production chain of firms the patent holder chooses to license. If the patent holder attempts to charge “too high” of a royalty

29. In a vertical production chain, one firm hands an intermediate product off to another firm to continue the production process, working in this fashion until the final good reaches the consumer. For example, glass lenses are frequently made by one firm, while another firm polishes them and inserts them into frames; the finished lenses are then distributed through a third link in the chain: specialty stores or optometrists.

30. For a discussion of “derived demand,” see PAUL A. SAMUELSON & WILLIAM D. NORDHAUS, *ECONOMICS* 213–14 (McGraw-Hill 1998).

rate to an intermediate goods supplier, that supplier will pass on its increased costs to its customers, who will then pass their increased costs on in the price of the final good. End consumers, thus, will face higher prices. As fundamental economics teaches, when prices rise for consumers, the quantity they demand generally falls. Therefore, in setting royalty rates, a patent holder must strike a balance, just as any price setter must. While raising the royalty rate increases the revenue dollars the patent holder earns on each covered unit sold, that royalty rate represents a cost to the licensees, who will pass the cost along the production chain to the end consumers. The higher price, therefore, will likely lower the number of units sold downstream and, in the process, lower the royalty base charged by the patent holder. If the number of units sold fall by more than the per unit revenue increases, the patent holder's profits will decline, making the royalty increase unprofitable.

These licensing dynamics are highly relevant to any discussion of first sale or patent exhaustion. When faced with a multilevel production chain, an IPR holder will set licensing rates to maximize its overall profits in light of the (anticipated or known) demand in the ultimate downstream market. This is true whether the IPR holder licenses one firm, or many firms, along the production chain. Even if an IPR holder charges a royalty at two or more points in the production chain, there is only one ultimate downstream market constraining the royalty rates charged.

Suppose that a patent holder determines that charging the downstream firm a royalty of five percent of the sales price maintains a healthy marketplace with sufficient demand so as to maximize licensing profits. Later, the patent holder decides to expand its licensing program to include upstream producers who provide inputs used by the downstream firms. Next, suppose that the patent holder keeps the downstream royalty rate at five percent and imposes an additional one percent royalty rate on upstream firms' sales. The upstream firms will face increased costs that will be passed on to their customers, the downstream producers, by raising their wholesale price to recoup the one percent royalty rate. Since the upstream royalty is passed on to the downstream firms by way of the wholesale price, it is as if the downstream firms faced a higher royalty rate, in

excess of five percent. The downstream royalty will thus exceed the optimal level for maximizing the patent holder's licensing profits and, by definition, the number of units sold will fall by more than the per unit profit increases. When charging an upstream royalty as well as a downstream one, the patent holder will therefore want to lower the royalty rate charged downstream. Adjusting both rates so as to keep the aggregate royalty earnings unchanged enables the patent holder to keep royalty profits at the optimal level. Thus, we did not need to impose the condition that aggregate royalties remained unchanged because that condition results naturally from the parties' profit motives in this ideal world with no market frictions.

With copyright and fine art sales, the "layers" at issue are typically not vertical ones in the production process, but rather different buyers separated by time. Generally, we do not have one business with IPR licensing multiple other businesses, but rather a rights holder licensing multiple end users, changing the dynamics considerably. The issue therefore becomes the impact of the IPR holder's ability to charge fees both on the initial sale and on a subsequent resale of the covered work.

Just as with patent holders, in our idealized world of perfect information, under a DDS regime, the first purchaser of a covered work would anticipate the application of royalty fees on any later sale, meaning that the first sale price would simply adjust to reflect the cost imposed on the resale. Therefore, as long as the rules were clear to all purchasers, as guaranteed by our benchmark assumption of perfect information, then the total payment that the creator received would be unchanged, just as the case with patents.

The key implication of the above analysis is that when frictions are not present, patent exhaustion and prohibitions on licensing multiple levels of the production-use chain cannot be motivated by the need for a mechanism to constrain the fees charged by IPR holders, as suggested in some of the *Quanta* filings.³¹ Indeed, at least when frictions and transaction costs are not an issue, the ability to charge

31. See Brief of Minebea Co. Ltd. as Amicus Curiae in Support of Petitioners, *supra* note 9, at 2–3; Brief for The United States as Amicus Curiae at 17–18, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937).

multiple parties in the production-use chain does not affect an IPR holder's income. Therefore, the next questions are whether and how the presence of transaction costs and market frictions might alter the above analysis.

III. INTRODUCING TRANSACTION COSTS

In most instances, of course, market frictions and transaction costs do exist and can be sufficiently large to affect the parties' decision making. While the idealized world described above is helpful in understanding key market dynamics, in order to make the economic assessment of the first sale doctrine meaningful, one must define a more realistic scenario. This more realistic scenario is one that recognizes the ubiquitous problems of transaction costs and marketplace frictions. Therefore, my analysis turns to the likely impact of such frictions on the optimal allocation of royalties among potential licensees along a given production-use chain, and examines the resulting implications for the first sale doctrine. In particular, I consider two important frictions: information frictions, such as the existence of private information or uncertainty, and cost pass-through frictions, where constraints prevent upstream licensees from fully passing through royalty cost increases to their downstream customers. I also discuss the additional precautions required when consumers are involved, as is the case in copyright and DDS sales.

A. The Importance of Information Frictions in Multi-Level Licensing

Consider first the implications of private information. Typically, royalty payments depend on the amount of the covered good actually sold in the marketplace, with a running royalty rate applied to a revenue royalty base. For example, suppose the parties agree to a royalty of five percent applied to the wholesale price of the covered good, payable each quarter. The licensee must then report to the patent holder its quarterly sales turnover (the total number of units sold, multiplied by the wholesale price). The literature has proposed many explanations for the predominance of licensing contracts of this sort, mainly related to private

information and uncertainty over downstream demand.³² As opposed to the sale of a physical good, where total sales can be estimated from the units of the input transferred, patents and copyright can allow for an unlimited number of units. That means IPR holders must be able to verify the quantities of covered goods sold by a licensee in order to calculate the royalty payments owed. It also means that licensees can have incentives to underreport sales in order to reduce their licensing payments. In this case, enforcing a contract is more complicated, requiring licensor monitoring and verification procedures.

In particular, if a patent license contract includes a percentage royalty rate or a per-unit fee—terms that are used commonly in patent licensing as risk sharing mechanisms³³—then the patent holder has a strong interest in setting the base for the royalty calculations on observable or verifiable quantities that licensed firms sell. Ambiguity over the relevant quantities sold, and hence over the basis for royalty payments, is an important difference between the licensing of intellectual property and the sale of a physical input. In the bike pedal example, if the designer were also the only component manufacturer, the quantities of bikes sold in the

32. For example, Alan W. Beggs, *The Licensing of Patents under Asymmetric Information*, 10 INT'L J. INDUS. ORG. 171 (1992) and I. MACHO-STADLER & D. PEREZ-CASTILLO, AN INTRODUCTION TO THE ECONOMICS OF INFORMATION: INCENTIVES AND CONTRACTS (Oxford University Press 2001) study models where the innovator uses a combination of royalties and fees to separate licensees with different cost reductions. Other papers show that royalties can be a signal of the quality of the innovation: Nancy T. Gallini & Brian D. Wright, *Technology Transfer under Asymmetric Information*, 21 RAND J. ECON. 147 (1990); or that royalties can be a response to the existence of moral hazard: Jay Pil Choi, *Technology Transfer with Moral Hazard*, 19 INT'L J. INDUS. ORG. 249 (2001) and Richard Jensen & Marie Thursby, *Proofs and Prototypes for Sale: the Licensing of University Inventions*, 91 AM. ECON. REV. 240 (2001). The specification of royalties can also be dictated by risk-sharing purposes. See Alain Bouquet, Helmuth Cremer, Marc Ivaldi & Michel Wolkowicz, *Risk Sharing in Licensing*, 16 INT'L J. INDUS. ORG. 535 (1998).

33. See, e.g., Morton I. Kamien & Yair Tauman, *Fees Versus Royalties and the Private Value of a Patent*, 101 Q. J. ECON. 471 (1986); Morton I. Kamien & Yair Tauman, *Patent Licensing: The Inside Story*, 70 THE MANCHESTER SCH. 7 (2002); Josh Lerner & Anne Layne-Farrar, *Valuing Patents for Licensing: A Practical Survey of the Literature* (Working Paper 2006), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1440292. Note that fixed per-unit dollar amounts can lower a licensee's cost uncertainty in the short run, whereas in markets where prices tend to fall over time, in the longer term such licensing schemes tend to lead to royalty rates that are disproportionately high in relation to the end product price.

final market would be easily inferred from the number of pedals sold in the wholesale component market. Once the pedal designer chooses not to manufacture the pedal itself, but instead to license its intellectual property to an upstream (pedal) manufacturer, the designer loses control and visibility over the number of units sold.

When private information is held by licensees, such as the actual number of goods sold incorporating the IPR, the IPR holder may want the flexibility to contract with multiple layers in the production chain. The IPR holder could then charge each link a partial fee that aggregates into an equivalent amount to the one that would have maintained had only one production level been licensed. Several factors support such licensing flexibility: monitoring costs, license enforcement costs, licensee incentives for under-reporting, and uncertainty regarding demand.

Monitoring is an issue when the relevant quantity sold is not easy for an IPR holder to verify. Initially, the IPR holder must decide the particular products to monitor. Frequently, it is difficult to identify the exact goods that a patent license covers. Manufacturers in high technology sectors often sell a wide variety of similar yet different products. For example, goods that incorporate multiple semiconductor chips tend to rely on hundreds, if not thousands, of patents.³⁴ For copyright, software provides a good example. Most commercial programs contain millions of lines of code, so determining whether a particular program infringes on another software producer's copyrighted code is quite difficult. In cases of this sort, determining the products that do and the products that do not incorporate the licensed IP can be difficult; furthermore, this determination is often subjective.

Once the products dependent on the IPR are determined, the IPR holder needs to monitor the sale of those products. Even with established licensing relationships, for many technology products that are shipped globally, the sale of intermediate components is difficult to monitor. In fact, due

34. See Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, in INNOVATION POL'Y & ECON. 1, 121 n.3 (Adam Jaffe, Josh Lerner & Scott Stern eds., MIT Press 2001) ("Nearly 5000 patents were granted in the U.S. in a recent single year, 1998, relating to 'microprocessors' alone, not to mention semiconductors more broadly.").

to the difficulty of monitoring, many firms have implemented licensing tracking systems for individual units (or boxes) being sold and shipped. For example, one firm instituted a special license aimed directly at the unreported sales problem: under this contract, a box containing components relying on the patented technology can only be shipped if proof of the license is displayed on the outside of the box.³⁵ Monitoring costs also explain the blanket licensing for copyrighted music, since determining the various licenses covering particular songs played by a radio station, and how many times those songs are played, would require exorbitantly expensive monitoring.³⁶

If monitoring licensee sales is costly for IPR holders, it makes sense to concentrate the monitoring effort on the production level, where this cost is lowest. The cost of monitoring a particular level of production might be different depending on the number of firms and the level of competition in the marketplace, or the closeness of the production stage to the final consumer. In particular, in markets where upstream prices are obtained as the result of private negotiations, but downstream prices are posted and publicly available, monitoring will be easier in the last stage of production.³⁷

Monitoring costs suggest that even if IPR holders license only one level of the production process, they should have the freedom to select the best level. On the other hand, when multiple production layers exist, licensing more than one level can provide several check points for verifying quantities sold, yielding the IPR holder with improved information on the downstream marketplace to correct for problems of

35. The per-batch licensing system was introduced by Philips (dubbing the program VEEZA) in place of its previous CD-R Disc Patent License Agreements. With VEEZA, a separate license is obtained for each shipment. The shipments are marked with a unique code signaling to the traders and retailers that the merchandise is licensed. See *Philips Intellectual Property & Standards, Licensing*, IP.PHILIPS.COM, <https://www.ip.philips.com/services/?module=IpsLicenseProgram&command=View&id=20&part=7> (last visited Feb. 21, 2010).

36. William M. Landes, *Optimal Sanctions for Antitrust Violations*, 50 U. CHI. L. REV. 662, 662 (1983).

37. For instance, a patent holder with patents on gasoline processing procedures can rely on the publicly posted prices for gasoline in the energy market. In contrast, a patent holder with a patent on, say, a drug delivery patch will not have reference to any public price lists for the delivery mechanism. In this case, only the final product price will be observable.

private information. In these cases, splitting fees across production layers can allow an IPR holder to obtain additional estimates of the quantity sold that complement the direct observation gathered through the level chosen for monitoring.

Likewise, license enforcement costs can affect an IPR holder's decision of which production levels to license. If the cost of enforcing a license differs across levels of the production process, the patent holder may prefer to shift or spread the allocation of the royalty burden. In case of a dispute, say because the licensee refuses to fulfill its payment obligations, the patent holder has different leverage depending on the characteristics of the licensee. Injunctions might be more effective against those firms whose product that incorporates the patent is highly profitable. This threat alone might be enough to allow the proper enforcement of the licensing contract. As a result, when enforcement is an important concern, the patent holder may want to shift the royalty burden towards those stages of production where competition is weaker and the licensee's profit margin is higher. In technology markets, these high-margin stages are often the ones closest to the final consumer, where product differentiation and brand reputation can make competition less fierce.

Licensee incentives to underreport product sales, thereby shrinking their royalty base and lowering their royalty payments, also can push patent holders to license at multiple stages of the production chain. This follows because the lower individual rates that result when multiple layers are licensed can reduce licensees' incentives to underreport the payments owed: less money is at stake so the incentive to misrepresent sales is lower. Recall from the idealized world discussion above that cost pass-through constrains the aggregate royalty payments the IPR holder receives. Frictions in cost pass-through are discussed in more detail below, but suffice it to say here that as long as *some* pass-through of royalty costs is likely, charging multiple levels implies lower royalty rates for any given level than if only one production level is licensed. Lower royalty rates at each stage can increase the odds of obtaining accurate information on the quantity of goods sold downstream (i.e., reduce the incentives for licensees to underreport).

Finally, information uncertainty can play a role as well. For example, placing the entire royalty burden on the first level of the production chain may be infeasible if it is difficult to forecast the ultimate downstream market demand at the time the upstream firm's license needs to be signed.³⁸ In those situations, the upstream manufacturer will have difficulty determining the appropriate wholesale price so as to pass on the correct royalty burden to the next level in the production chain. Splitting the royalty burden across production levels can help to spread the risk of demand uncertainty more evenly across market players. Moreover, timing might matter as well if negotiations with downstream producers, selling goods into the end market, can take place at a later date. At that point, the downstream firm better knows its needs, whereas at the time of negotiations with the upstream producer the downstream production needs would have to be estimated.

The difficulties introduced by uncertainty and private information therefore suggest a positive role for multiple-level patent licensing under certain circumstances. In those cases, the ability to relax the strict application of the first sale doctrine can lead to more efficient contracts by providing the patent holder with additional information on downstream sales or with easier contract enforcement.

Thus, while the first sale doctrine played no role, positive or negative, in the ideal frictionless world that this paper initially considered, in the presence of information frictions, a strict application of the first sale doctrine to patent licensing across production levels actually can be harmful. Some frictions suggest that single-level licensing is best (e.g., finding the lowest cost monitoring environment or the marketplace that best facilitates license enforcement), but other frictions indicate that multiple level licensing can be quite important (e.g., collecting royalty base information from multiple points to improve overall demand information or lowering licensees' incentives to underreport relevant sales). Accordingly, preventing firms within a production chain from

38. As LG Electronics argued in the case, "petitioners' [Quanta's] approach would demand that the full and final value of these patents be determined at a single point in the distribution chain, where the relevant information simply does not exist." Brief of Respondent at 32, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937).

contracting around patent exhaustion in circumstances where it makes sense to do so will increase the likelihood of underreporting for certain patent holders. As a result, information frictions provide an economic justification for flexibility in contracting around patent exhaustion (before the final good is sold to the end consumer).

B. Extending the Logic Beyond Patents

As already noted, the production of copyrighted goods and artwork often does not involve multiple production layers. What, then, does the introduction of information frictions imply for relaxing the first sale doctrine for these IPRs?

For digital copyright, some of the most important information frictions relate to the monitoring of copies and the enforcement of the license contract. In fact, the prevalence of online piracy has been a key motivation in the creation of digital rights management schemes.³⁹ Recall that end buyers of online copyrighted works do not face the transaction costs that offline end buyers do (such as the costs associated with creating and distributing physical copies). As a result, proponents of DRM argue that the rights themselves must be adjusted because otherwise the incentive structure created by the grant of a copyright will be completely eviscerated by the technological realities of today's marketplace.⁴⁰ Hence, imposition of restrictions that affect first sale—such as the imposition of time limits on the use of a copyrighted work and the prevention of any resale of the work—can be justified, under this view, as resetting the original balance of seller-buyer rights achieved through copyright law offline.⁴¹

On the other side of the argument, there is the potential for very different degrees of knowledge, information asymmetry, and understanding between copyright sellers and

39. See Bach, *supra* note 6.

40. See CONGRESSIONAL BUDGET OFFICE, *supra* note 4, at 34–35.

41. Of course, not all DRM claw-backs can be justified on these grounds. Just as technology permits copyright buyers to share copies widely, enabling friends to avoid the purchase price, so too does technology lead to abuses by rights holders who impose restrictions simply because they can. For a discussion of the abuses of DRM, such as using DRM to violate private rights, see Richard Watt, *The Past and the Future of the Economics of Copyright*, 1 REV. ECON. RES. COPYRIGHT ISSUES 151, 164 (2004).

buyers. Unlike the above discussion of multilevel patent licensing, where a business entity is on each side of the negotiation table, oftentimes copyright transactions are business-to-consumer. For negotiations over a blanket license to play a catalog of music between ASCAP (American Society of Composers, Authors and Publishers) and a large department store chain, for instance, we do not need to worry as much about the disparate sophistication of the parties. On the other hand, disparate sophistication becomes relevant when considering a consumer's purchase of an e-book or downloadable song.

Concern regarding certainty over the end price and the rights conveyed to consumers are likely to give us greater pause when considering whether first sale can be relaxed contractually for the sale of copyrighted goods. But even in the real world where information frictions are an everyday issue, we may still find some, albeit relatively narrow, situations where allowing for such relaxation makes sense. Consider the sale of software and music over the internet. Over the past few decades, so-called "click wrap" licenses⁴² have become quite common, given how easily sellers can now impose them.⁴³ It might be reasonable to assume that consumers are now familiar enough with these forms of contracts to allow copyright holders to sell certain restricted-rights products, like e-books that can be read only on one form of digital reader and cannot be copied to multiple devices.

With click wrap licensing, if the copyright holder provides adequate notice of the restriction on rights that end users would not otherwise expect, consumers may actually benefit from the restriction by receiving more information about the grant of rights. In particular, in this instance contractual flexibility regarding first sale doctrine could provide consumers with a broader range of available products

42. A click wrap license is an online license where the terms and conditions of the license require the purchaser to click on an "I agree" or an "I accept these terms and conditions" button before consummating an online sale. See, e.g., Francis M. Brono & Jonathan A. Friedman, *Maximizing the Enforceability of Click-Wrap Agreements*, 4 J. TECH. L. & POL'Y 3 (1999).

43. Patrick J. Mondy, *I Accept the Terms in the Agreement: Market Efficiency in Clickwrap Agreements and Open Source Software*, 2 SEVENTH CIRCUIT REV. 540, 546 (2007).

at a broader range of prices.⁴⁴ If consumers can make an informed tradeoff between more rights (the ability to copy the work to another personal device, for instance) for more money, overall consumer welfare could be higher. Consumers with marginal valuations for the right to make digital copies for other devices may purchase an online copyrighted good with restricted rights that is offered at a discount when they would not have purchased the same copyrighted product with fuller rights for a higher price. In this case, the ability to impose restrictions broadens consumer choice in the overall market.

A similar careful regard toward end users, and their ability to learn of and understand the rights conveyed by the seller, would apply in regards to works of art for the DDS rule. The most relevant of the information frictions in this instance is primarily mutual uncertainty, rather than any privately held knowledge by one or the other party. With new artistic works, paintings, sculptures, etc., it will often be the case that neither party knows how strong the ultimate demand for the artist's work will be at some later point in time when the current work may be resold. In fact, many works of art never become popular enough to have a resale market at all.⁴⁵ The ultimate popularity of an artist's works is also influenced by that artist's actions. Indeed, the mere act of making multiple works that are similar to one another in form can lower the price of any one work. Moreover, it is likely that even in those jurisdictions where a DDS rule is in force, it is irrelevant for the vast majority of artwork, as the pieces will never be resold or the DDS rule will not be actively enforced.

Uncertainty over ultimate popularity was a particular motivation for the introduction of the DDS rule over works of art.⁴⁶ At the time of the initial sale, an artist might be young

44. See *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (1996).

45. See John Solow, *An Economic Analysis of the Droit de Suite*, 22 J. CULTURAL ECON. 209, 210 (1998).

46. *Id.* at 211.

[This paper] focuses on the decision to produce at different points in time; the resale price of a piece of art will depend on the prices of substitutes and complements including, in all likelihood, later works by the artist. The key insight is that the residual interest in early works provided by the resale royalty gives the artist an incentive to maintain the value of those works that is absent without the royalty.

and relatively unknown; when the work is resold at a later point in time, the market value of the work is more likely to be established. Hence, a regime where the creator receives an additional royalty payment at the time of resale can provide better incentives for the creation of new works in the first instance, can create incentives for an artist to cultivate and maintain demand for his or her creations, can moderate the production of near-copies, and can better ensure an economically appropriate return for the IPR holder.

In the creation of these incentives, DDS is analogous to the use of royalties for patent licensing. In particular, splitting the total payment for a given work between today's price and tomorrow's (potential) royalty allows the artist and the buyer to share some of the risk involved in trading works of art. If the buyer anticipates little to no aftermarket for an artist's work, then no royalty payments will be expected and the current price will reflect the full value assessed. If, on the other hand, the buyer expects the work of art to be popular in the future, to be eligible for resale, and anticipates that DDS will be enforced, then the buyer will pay a lower price today reflecting the present value of the later royalty payment. If the work is more popular than anticipated, the artist will receive a windfall in actual royalty payments exceeding their forecasted amount at the time of the original sale. If, on the other hand, resale values do not materialize as anticipated, the artist will be undercompensated because later royalty payments will fall short of the original price offset. To the extent that an artist can control (or more realistically, contribute to) the future demand for her own works, say by giving more shows or promoting her work at galleries and on the Internet, DDS may help to align incentives and give artists financial reasons for making such efforts at promotion.⁴⁷

However, any positive forces for risk sharing and incentives to maintain market value would have to be weighed against the costs imposed on buyers of artwork. As with copyright, art buyers would need to fully understand the

Id.

47. *Id.* at 210 (citing Larry Karp & Jeffrey Perloff, *Legal Requirements that Artists Receive Resale Royalties*, 13 INT'L. REV. L. & ECON. 163 (2003)).

DDS legal regime regulating the sale of artworks in order to properly assess prices, of both today and tomorrow.

C. The Importance of Cost Pass-Through Frictions in Downstream Licensing

The discussion, thus far, has highlighted the important role that information plays in efficient licensing. We have largely assumed that the various manufacturing firms pass costs through to their customers in setting wholesale prices, providing a link between upstream and downstream markets. How does the analysis change if frictions in the marketplace limit cost pass-through? What if technological or institutional constraints prevent wholesale prices from fully adjusting to changes in the royalties?

Consider an extreme case where the upstream manufacturer has no control over price setting at all. In this environment, the wholesale component price will be exogenously determined and no pass-through will be possible. Thus, when the IPR holder implements different ways to split the total royalty, the allocations could affect end markets and IPR royalty revenue could differ in the aggregate as well as in the components. Zero cost pass-through on the part of manufacturers is, however, just as unrealistic as the absence of all frictions or transaction costs. More realistically, component manufacturers facing pricing frictions will be able to pass some, but not all, cost increases through to their customers.⁴⁸

In the case where cost pass-through is partial, suppose the patent holder attempts to charge the full royalty burden on the upstream component maker. The upstream manufacturer—knowing that it cannot recoup the royalty expense applicable downstream because it cannot adjust the wholesale price to fully reflect the cost increase resulting from the royalty charge—will negotiate with the IPR holder with its own profit margin in mind. Unless the licensee has no bargaining power at all (such as after irreversible investments are made), it is likely that the resulting royalty rate will equal (or be close to) the partial rate reflecting the value of the patent to the upstream production process that

48. See DENNIS CARLTON & JEFFREY PERLOFF, MODERN INDUSTRIAL ORGANIZATION 254 (Addison Wesley 2004).

would have resulted in the frictionless world with full pass-through. This means, once again, that the patent holder cannot double dip in multilevel production licensing. The IPR holder will therefore need to license the downstream manufacturer, as well as the upstream one, in order to obtain the full economic return on its IPR.

Note that the derived demand constraint discussed earlier may be limited when the upstream firm cannot fully pass through any royalty burden in the wholesale price. Restricted pass-through attenuates the link to end consumer demand, demand that previously acted to moderate the aggregate royalties assessed by the patent holder. The patent holder may be able, if it holds sufficient bargaining power (e.g., ex post licensing), to increase its total royalty earnings as compared to the frictionless full pass-through environment. Even with ex post licensing, however, the patent holder would still need to be mindful of a licensee's incentives to reduce its supply in the components market. In practice, this could lead to an equivalent result as full cost pass-through, or lead the component manufacturer to work around the patent, reducing royalty income to zero. Suppose, though, that neither of those constraints for the component supplier was substantial, so that the patent holder could indeed double dip to some extent, in the sense of increasing its aggregate royalty income by licensing multiple production levels as compared to licensing a single level.⁴⁹

Still, this scenario would not match the complaints raised during the *Quanta* case because those arguments hinged on pass-through that harmed end consumers. Here, on the other hand, as long as some pass-through to consumers is possible, the downstream royalty would be set with end consumer demand in mind, just as with single level licensing, because that is how the patent holder maximizes its royalty revenues on the downstream layer of production. The additional royalty income would derive from bargaining power in licensing the upstream firm as well as the downstream firm. However, the limited cost pass-through implies that any additional royalty income above the single level amount is a firm-to-firm transfer of profits from the upstream

49. As before, multilevel licensing is restricted to business-to-business negotiations and does not include end consumers.

manufacturer to the patent holder; consumers would not be harmed.⁵⁰ As it appears, in most instances patent holders are under-compensated.⁵¹ Even if a patent holder could earn more with multi-layer licensing, such a licensing scheme may still not translate into over-compensation. End consumers would only be harmed to the extent that end prices rise as a result of pass-through; but, if costs are passed through then we are back in the frictionless world where double dipping is not profit maximizing for patent holders, and thus not practiced.

Double dipping, in those instances where it is possible, is a special instance of patent holdup. It is difficult to see why the doctrine of patent exhaustion, however, is the best solution for dealing with this narrow subset of patent holdup. First, since consumer welfare is not affected, a clear case would need to be made that social welfare was harmed, say through detrimental effects on innovation or industry investment incentives, before any government or court intervention is warranted. Moreover, the harm identified from double dipping would need to outweigh the benefits of contracting around patent exhaustion, as discussed above. While we might have “fairness” objections to one firm having the ability to exploit another at the bargaining table, we should bear in mind that pricing exploitations happen with regularity throughout the economy. These exploitations are by no means limited to IPR licensing and other more general forms of law (contract, fair business practices, etc.) are available for dealing with the egregious cases. It is unclear what patent exhaustion has to offer that these other legal doctrines cannot deal with already, especially in light of the collateral damage that could occur from restricting firms’ ability to contract around exhaustion in business-to-business dealings.

50. Recall that any relaxation of patent exhaustion would apply only in firm-to-firm negotiations and royalties would not be charged at the end consumer level. Note also, that even with limited pass-through of costs, upstream firms will not agree to a royalty in excess of the value they receive unless the license negotiation takes place *ex post*, when irreversible investments can be exploited, or when they have no bargaining power for some other reason. I discuss *ex post* licensing below. See *infra* Part IV.

51. See Vincenzo Denicolò, *Do Patents Over-Compensate Innovators?*, 22 ECON. POL’Y 679 (2007).

As another example of cost pass-through frictions, consider the case where a single upstream producer sells to several downstream firms in different market niches such that the same component input has a different added value. This would be the case, for example, if the same semiconductor chip were incorporated in mobile phones, external USB modems, and certain laptop computers. In this case, complete cost pass-through would mean that the profits for all parties would be the same regardless of whether the patent holder charges a royalty upstream and (possibly) different royalties downstream, or if it charges the full royalty upstream and the upstream producer modifies the price of the intermediate product appropriately to pass the royalty cost along to each (differentiated) downstream producer.

Suppose once again, however, due to arbitrage or antitrust cautions, that the upstream component manufacturer is unable to pass on all of its costs. In this case, the strict application of the exhaustion doctrine could make it difficult or impossible for the patent holder to achieve the economically justified return on its IPR, or to ensure licensing efficiency. Suppose the patent holder only licenses the upstream firm. If the patent holder charges the royalty dictated by the highest value use of its IPR (say the mobile phone handsets in the example above), that rate will be too high for the component maker to make sales to any lower valuation uses (like USB modems). In this case, the component maker will choose not to serve the lower value downstream market at all, and will instead opt to keep royalty costs at a reasonable level so as to maintain overall profits. Alternatively, if the patent holder sets the royalty rate according to the lowest value use of the IPR, it is simply transferring its profits to the component maker, who will pocket the difference on any sales made to the higher value users. Hence, if licensing is to occur at the upstream level, the patent holder must be able to discriminate amongst the downstream production uses. This requires the ability to restrict uses of the IPR and requires accurate reporting of the division of uses on the part of the licensee. If, due to monitoring imperfections, such discriminatory pricing is not workable, the patent holder will be under-compensated, reducing its incentives to license in the first place, or lower

value downstream markets will not be served, to the detriment of end consumers.

The way around these problems is for the patent holder to license the downstream markets directly, ignoring the upstream level. In this case, royalty rates can be tailored to the corresponding valuation of the IPR in any given downstream market. While this might appear to be an easy solution, preserving the exhaustion doctrine and avoiding problems that stem from an inability of upstream producers to fully pass costs through to the next level of the production chain, that conclusion relies on the absence of any other frictions. As discussed earlier, information asymmetries may indicate licensing upstream is the best approach, but cost pass-through limits may prevent that. In other words, when both kinds of frictions are present, multiple production-level licensing may be the best solution to minimize all concerns.

Similar forces may be at play in copyright licensing, although these cases are less frequent because there are fewer instances of multilevel production relevant for copyright. As one example, it is reasonable to expect that some software modules will find multiple applications for embedding in products downstream. In this case, where the buyer and seller are both businesses, it can make sense for copyright holders to limit resale—for example, to prevent arbitrage from a low value/low licensing fee market toward a higher value/higher licensing fee market.

The fundamental result is that in the face of transaction costs, it can be important to allow IPR holders the flexibility to license multiple parties and place certain contractual restrictions on the buyer's rights. Charging just one versus charging multiple parties is not the pivotal element for social welfare. In fact, in the presence of marketplace frictions, the way the total licensing burden is split among (business-to-business) licensees is likely to reflect the cheapest and most convenient way to implement licensing, and this split is bound to differ across firms, industries, and sectors of the economy. In other words, charging multiple firm levels might be crucial to maintaining both economically justified rewards and efficient licensing.

Thus, the analysis presented here suggests that, in many circumstances, strictly applying the patent exhaustion doctrine in business exchanges could create economic

ineficiencias. Moreover, in some circumstances with copyright and DDS, relaxing first sale can actually be beneficial to end buyers. While it is clear that one of the benefits stemming from the patent exhaustion or first sale doctrine is the certainty that later users of a product will not be sought after for a license, if this was an important risk in business-to-business transactions it could easily be addressed by using first sale as the presumed or “default” (but not mandatory) rule. In that case, first sale would be overturned through explicit clauses of IPR licensing, giving proper notice to buyers, when other concerns were of greater importance. If a patent license contract to a firm did not specify limitations or any subset of rights that are not passed through to the licensee’s customers, then first sale would apply; otherwise, it would not. For other IPR, end user notice would need to be clear and accessible, and the rights holder would need to provide an economic rationale for the restriction (such as offering lower priced products to expand consumer choice) to ensure that the practice was not abusive. As explored above, it is certainly possible that such justifications could be made.

IV. COMPETITION CONCERNS RELATED TO MULTI-PARTY LICENSING

Based on a reading of the various amici briefs in the *Quanta* case,⁵² the primary competition concern at issue with the patent exhaustion doctrine appears to be double dipping, or “excessive” royalty earnings by the IPR holder. Other potential concerns that emerge from the analysis presented in this paper are: (1) “double marginalization,” where the presence of multiple levels of production increases end-market prices above even the monopolistic level; (2) “royalty stacking,” where multiple patent licenses can stack upon one another to create a very high aggregate licensing cost for manufacturers; and (3) raising rivals costs, where a vertically integrated IPR holder could use licensing rates to reduce

52. Brief of Dell Inc., Hewlett-Packard, Co. & Gateway, Inc. as Amici Curiae In Support of Petitioners at 12, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 553 U.S. 617 (2008) (No. 06-937) (“Allowing the patent owner to multiply its recovery by extracting a new, duplicative royalty at each stage of ownership does not promote the progress of science and the useful arts and thus does not serve the purposes of the patent system.”).

downstream competition. In this section, I consider each of these concerns in turn.

A. Double Dipping

As explained above, in a frictionless world, double dipping is not possible in multilevel licensing; intermediate firms will simply adjust wholesale prices to pass on any royalty costs, so that end-market suppliers bear the full royalty burden whether it is charged to them directly or not. In a world with limits on the ability of intermediate manufacturers to pass through costs, some double dipping may be possible, but in that case it represents a transfer between firms, not a transfer from consumers to producers. Thus, the concern is not realized.

The key factor providing the avoidance of double dipping is the ability of the parties to contract *ex ante*, before potential licensees have made any sunk investments. During *ex post* negotiations, if such irreversible investments have been made, then multiple licenses may indeed double dip, but a better name for this behavior is patent hold up. In other words, as already noted, double dipping is really only a special, rather narrow, case of *ex post* extortion that is only possible when downstream firms have a good in the market before a license on the IPR that good relies upon is agreed. The patent exhaustion doctrine is irrelevant in this instance: irreversible investments can be exploited at one level of production or at many levels. Thus, a strict application of exhaustion is not an effective means of preventing such opportunistic licensing, as the ability to hold up licensees hinges on the presence of irreversible investments, *ex post* negotiations, and asymmetric bargaining power, and does not rely on an ability to license multiple parties in the production chain.

A variant of double dipping has also been expressed as a concern over DDS rules. For example, some authors have raised the theoretical concern that DDS rules might reward authors in the phase of their careers when they are relatively less productive and might also distort the market by diverting artworks from DDS-friendly countries towards those jurisdictions where artists do not receive such additional

royalties.⁵³ In contrast, applying the analysis presented here to DDS rules implies that if uncertainty is not large, the prices initially paid to an artist would adjust to reflect expected DDS payments made in the future. Moreover, as explained above, DDS may increase an artist's incentives to maintain the market value of his works.

In the face of dueling theories, the matter must be resolved empirically, and quantitative studies of DDS rules indicate that the ex post royalty does not create inefficiencies. While, short of a large-scale survey, there is no way of knowing whether DDS does or does not improve artists' incentives to maintain the future value of their works, the evidence presented in economic literature shows that the impact on future prices for artwork covered by DDS is negligible, as are any trading volume effects.⁵⁴ If anything, the quantitative studies indicate that trading volume increases in DDS-friendly countries as compared to countries without DDS rules. This makes sense if DDS works to motivate artists to expend effort to maintain their reputations and the value of their works of art over the artist's lifetime.⁵⁵

B. Double Marginalization

The other "double" problem that emerges in the analysis above is double marginalization.⁵⁶ This issue arises when firms at multiple levels of the production process each want to charge a profit margin on their portion of the sale. As the good works its way toward the final market the margins stack over production levels, raising end consumer prices. In fact, industries characterized by double marginalization tend to have prices that exceed the monopoly price level.

One might suppose that the presence of double marginalization would suggest that the IPR holder charges all of the royalty downstream, lest it increase margins at both

53. See Ginsburgh, *supra* note 9, at 68–69.

54. See, e.g., Kathryn Graddy & Chanont Banternghansa, *The Impact of the Droit de Suite in the UK: An Empirical Analysis* 1 (CEPR Discussion Paper No. DP7136, 2009).

55. Or, alternatively, these results are as expected if we believe that DDS rules are not actively enforced.

56. For a discussion of the economics of double marginalization, see W. KIP VISCUSI, JOHN M. VERNON & JOSEPH E. HARRINGTON, JR., *ECONOMICS OF REGULATION AND ANTITRUST*, 221–23 (MIT Press 2000).

production levels and hence exacerbate double marginalization. As it turns out, however, this conjecture is not accurate, since with any degree of cost pass-through the intermediate wholesale price charged by upstream manufacturers will adjust as the upstream royalty rate changes. So, just as with double dipping, double marginalization distortions are unrelated to the first sale doctrine. Instead, such problems derive solely from the presence of various production levels in an industry, and from the relationship between the firms active at those levels.⁵⁷

Note that patent licensing could actually undo the problems of double marginalization if the IPR holder were able to charge an unorthodox two-part tariff consisting of a negative royalty rate and an upfront fixed fee. Specifically, an IPR holder could extract an economically justifiable return through the imposition of a fixed license fee that does not affect any of the producer firms' marginal quantity decisions.⁵⁸ The IPR holder could then undo the double marginalization arising from the relationship between the upstream and the downstream producers by charging a *negative* royalty rate upstream.⁵⁹ If this were possible, it would lead the downstream firm to charge the monopoly price, a price lower than that under double marginalization. By lowering the downstream producer's costs (and hence its prices), this licensing regime would result in higher downstream quantities sold. In this case, the upstream producer would be compensated for its inability to charge a profit margin through the receipt of the negative royalty rate transfer payment from the IPR holder. The IPR holder would be compensated for the transfer to the upstream producer as

57. With vertical integration, one entity controls the full production process. Even if subsidiaries and affiliates are involved, there is still some level of centralized control that will internalize pricing issues.

58. Indeed, early economic analysis predicted that all or most patent licenses should involve lump sum payments so as not to distort licensee's marginal production decisions. In the absence of any uncertainty, such payments are more economically efficient. See, e.g., Morton I. Kamien, *Patent Licensing*, in HANDBOOK OF GAME THEORY WITH ECONOMIC APPLICATIONS 347 (Robert J. Aumann & Sergio Hart eds., 1992) ("[T]he patentee's licensing profits are higher if he auctions licenses than if he employs a royalty, except if the invention is drastic or the industry is perfectly competitive.").

59. That is, the patent holder would charge a "normal" royalty to the downstream producer, but would pay the upstream producer (transferring funds from the downstream producer) rather than charging it a royalty.

well as the value of its IPR through the fixed fees charged downstream. While this approach is theoretically possible, the ability to charge negative royalty rates seems highly unrealistic at best, making this option little more than a hypothetical thought exercise.

C. Royalty Stacking is a Separate Issue

Yet another complaint frequently raised by patent licensees is the issue of royalty stacking.⁶⁰ While double marginalization involves the “stacking” of royalties over vertical layers of production, under royalty stacking, the concern is a horizontal one: the presence of multiple patent holders that all charge royalties to the same licensee. Clearly, the problem of royalty stacking does not necessarily involve licenses at multiple production levels. Nonetheless, royalty stacking might still be relevant for patent exhaustion in multilevel production industries.

Consider an industry with two production levels and numerous upstream patent holders. Just as in one of the scenarios discussed above, the many patent holders could license just the upstream producers, who would then pass through full rights to the downstream firms to complete the good’s production for sale to end users. Alternatively, the patent holders could split the rights across levels, charging a lower rate to the upstream firms, but not allowing the pass through of full rights so the downstream firms would need licenses as well. Downstream firms could have one of two preferences in this scenario. They might prefer that the upstream firms acquire full pass through rights, so the downstream firms do not need to directly pay any licensing fees. Upstream firms might reasonably charge a premium for this service, since they would take on all of the coordination costs and would be providing their customers with complete input cost certainty. This certainty is especially important when there is ambiguity over the set of patents reading on the downstream product. Alternatively, some downstream firms might prefer to pay a lower price to the wholesaler for just the input, acquiring any patent rights they might need through their own direct negotiations. If downstream firms

60. See, e.g., Mark A. Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEX L. REV. 1991 (2007).

believed they did not need licenses to the full set of upstream patents, or if they felt they were in a stronger bargaining position than the upstream component supplier, they might prefer this route. It would involve increased transaction costs, as the downstream firm would need to negotiate patent licenses on its own, but it may lower the aggregate royalty payments the firm had to make.

The scenario likely to be more prevalent in any given industry is an empirical matter that we have little information about, given the confidential nature of firms' patent negotiations. Here, the point is that allowing flexibility to contract around patent exhaustion allows an efficient market outcome to emerge, without courts needing to guess at the particular set of preferences firms may have. Thus, even in the presence of potential royalty stacking, it is economically efficient to allow for the option to contract around patent exhaustion.

D. Raising Rivals' Costs Does Not Depend On Multi-Level Licensing

The final potential competition concern relates to raising rivals' costs. Suppose that the upstream producer also offers a substitute good to the one sold by the downstream firm. It is a well-known concern that an integrated firm with an upstream presence might raise the price of its input component as a way of increasing the downstream firm's costs and hence its prices. This is a means for relaxing downstream competition and increasing the sale of the integrated suppliers' substitute product.⁶¹

While this dynamic might raise the wholesale price that the integrated firm charges for input components, the royalty rates charged by the IPR holder would not play any role in that decision. Even when the IPR holder charges a different combination of royalties up and downstream, with some degree of cost pass-through, the upstream manufacturer will respond by changing the wholesale price, leaving the total input cost for the final product roughly unchanged. As a

61. See, e.g., Timothy J. Brennan, *Understanding "Raising Rivals' Costs"*, 33 ANTITRUST BULL. 95 (1988); Steven C. Salop & David T. Scheffman, *Raising Rivals' Costs*, 73 AM. ECON. REV. 267 (1983); David S. Sibley & Dennis L. Weisman, *Raising Rivals' Costs: The Entry of an Upstream Monopolist Into Downstream Markets*, 10 INFO. ECON. & POL'Y 451 (1998).

result, the application of the first sale doctrine is not at issue in markets where upstream manufacturers are vertically integrated and there are a number of downstream competitors.

Alternatively, suppose that it is the IPR holder that is vertically integrated into downstream production. In this case, the royalty rate might be used as a vehicle for disguising a raising-rivals'-cost scheme.⁶² That being said, if the IPR holder has sufficient market power to raise the royalty rate to its downstream competitors in this fashion, presumably the patent holder could also simply raise its component wholesale price. The only instance where the royalty rate might offer the only vehicle for raising rivals' costs would occur when the intermediate components were supplied competitively while the IPR was not (i.e., the IPR had no viable substitutes). In this case, downstream producers could obtain their inputs from a rival upstream component supplier to the IPR holder. Under this circumstance, the patent holder would intentionally set the royalty rate above the level that was optimal for maximizing its IPR licensing profits in order to soften downstream competition in the goods market so as to earn a greater share of product sales for itself. In other words, the patent holder would set the royalty rate to maximize profits from the sale of goods, not from the licensing program in isolation. But in either of these cases, where the patent holder is vertically integrated, licensing multiple levels is not necessary to accomplish the raising-rivals'-costs strategy: the IPR holder can simply license the level of production where it directly competes, ignoring any other levels of production.

Thus, none of the above possible competition concerns within multiple level production turn on the application of the patent exhaustion doctrine. While double dipping, double marginalization, royalty stacking, and raising rivals' costs are all serious concerns with antitrust implications, none of them depend on an IPR holder's ability to charge royalties at multiple layers of the production chain. Instead, it is other factors, like the presence of irreversible investments that can

62. Because IPR is much harder to value than tangible goods, it will be more difficult for licensees and competition authorities to identify overpriced IPR meant to raise a rival's costs.

be exploited or the existence of vertical integration, that enable these opportunistic behaviors.

V. CONCLUSIONS

The analysis presented in this paper has focused on the economics of patent licensing in a complex industrial environment, with a less detailed examination of copyright licensing and artwork sales under DDS. In an idealized setting, with no transaction costs or asymmetries of information, application of the patent exhaustion doctrine (either strictly or with flexibility) within a multilevel production setting has no impact on social welfare. Essentially, it is irrelevant for setting aggregate royalties, for determining final market prices, or for the quantities sold. Charging just one (versus multiple) firms along the production chain is not the pivotal element for social welfare.

Considering an idealized world, with no frictions or transaction costs, is a useful thought experiment because it clarifies the relevant dynamic competition factors and provides a meaningful benchmark case. That being said, to understand real world licensing, one must add the real frictions and transaction costs into the analysis. For instance, many industries are likely to be characterized by private information, uncertainty over product demand, or wholesale input pricing constraints. Even in this more complicated and more realistic world, though, there is no justification for placing absolute restrictions on the ability of IPR holders to split fees among multiple production layers. In fact, in the face of transaction costs and frictions, a strict interpretation of the patent exhaustion doctrine is likely to generate welfare losses in the economically justified reward and efficiency dimensions of licensing.⁶³

In light of the realistic scenarios where the ability to contract around exhaustion is important, a flexible approach to the doctrine's application is warranted on the basis of economic efficiency. Certainly not all circumstances will call for multilevel (or across-time, for copyright) licensing. Indeed, single level licensing can be more economically efficient.⁶⁴ But, when circumstances call for multilevel

63. See *supra* Part III.B.

64. See *supra* text accompanying notes 49–51.

licensing—for instance, when royalty base monitoring can be improved by obtaining output reports from multiple production points or when sharing the royalty burden across production levels reduces licensees' incentives to underreport their relevant sales—firms should have the freedom to contract around first sale. This ability to contract around first sale must be paired with the certainty that a court will not undo the contract at some later point in time.

Certainly one of the key concerns raised in the *Quanta* case, that a patent holder charging royalties at multiple layers of the production process could be engaged in double dipping, appears to be misplaced. In actuality, this concern will only apply under very narrow circumstances, and it will not have a direct consumer welfare impact in any event. Certainly with business-to-business negotiations, the main motivation of providing pricing certainty appears to be far less important, as the parties are likely to be sophisticated and able to negotiate reasonable contracts. The primary message that emerges from the analysis presented here is, therefore, that a balanced approach in the application of the patent exhaustion doctrine is required. Expanding the strict application of the doctrine in multilevel production settings for licenses between two businesses would likely have detrimental effects for end consumers and industry participants. A more flexible application of the exhaustion doctrine, where its presence would be inferred only in the absence of license clauses to the contrary, is the preferable approach from an economic perspective.

In other settings involving end consumers, first sale has more of a role to play in creating certainty, but even here there are some narrow circumstances where relaxing the doctrine could be beneficial to consumers. Keeping the possible benefits to consumer welfare in mind, such as increased consumer choice and potentially lower prices for restricted rights, will lead to a more sensible application of the first sale doctrine.