Patent Licensing in Vertically Disaggregated Industries: the Royalty Allocation Neutrality Principle

Anne Layne-Farrar
Charles River Associates

Gerard Llobet
CEMFI

Jorge Padilla*
Compass Lexecon

7 May 2014

Abstract This paper investigates patent licensing in vertically disaggregated industries, where patent holders may license to upstream producers only, downstream producers only, or to both upstream and downstream producers. We consider whether consumer welfare will be greater if the patent holder's ability to license multiple parties along a production chain is restricted. We also analyse whether a policy that restricts licensing to upstream manufacturers constitutes appropriate public policy. These questions have significant policy implications. Under the legal doctrine of first sale, or patent exhaustion, a patent holder's ability to license multiple parties along a production chain is restricted. How and when such restrictions should be applied is a controversial issue, as evidenced by the US Supreme Court's granting certiorari in the *Quanta* case. Some commentators have even argued that refusing to license to upstream component manufacturers may constitute an abuse of dominance and thus infringe the competition laws. We find that under ideal circumstances how royalty rates are split along the production chain has no real consequence for social welfare. Even when we depart from ideal conditions, however, we still find no economic justification for restrictions of the patent holders' ability to license multiple parties or to license to downstream producers only.

*Anne Layne-Farrar is an economist at Charles River Associates; Gerard Llobet is a professor at CEMFI; and Jorge Padilla is an economist at Compass Lexecon and a Research Fellow at CEMFI. An earlier version of this paper was circulated under the title "An Economic Take on Patent Licensing: Understanding the Implications of the First Sale Patent Exhaustion Doctrine". We wish to thank Guillermo Caruana, Richard Epstein, Scott Kieff, and Klaus Schmidt for helpful comments and suggestions and Alina Marinova for research assistance. Financial support from Qualcomm is gratefully acknowledged. The ideas and opinions in this paper are exclusively our own. Comments should be sent to llobet@cemfi.es.

1. Introduction

This paper investigates patent licensing in vertically disaggregated industries, where patent holders may license to upstream producers only, downstream producers only, or to both upstream and downstream producers. We try to disentangle the impact on profits and consumer welfare of these alternative licensing strategies. Will consumer welfare be greater if the patent holder's ability to license multiple parties along a production chain is restricted? What are the likely welfare implications of a policy that mandates patent holders' to license to upstream manufacturers only, thus prohibiting them to license to downstream manufacturers, either in isolation or in combination with the upstream manufacturers?

These questions have significant policy implications. In the summer of 2008, the U.S. Supreme Court issued its opinion for *Quanta Computer et al. v. LG Electronics, Inc*.[1] The central issue of debate in this case was the doctrine of first sale patent exhaustion – whether, or more precisely when, the first sale of a product embodying patented technology by the patent holder limits in some way the patent rights, so that the patent holder cannot pursue any later purchasers for a license. Issues of double payment and upstream/downstream rights have been debated in ordinary patent cases as patentees have moved from licensing component manufacturers to licensing their downstream customers. Some commentators have said that in cases involving complex products with many components, calculating a royalty base that is broader than the value of the components covered by the patent may over-compensate patent holders when the patent at issue covers only some components and those components are not the sole drivers of consumer demand for the product.[2] According to this view, a policy that restricts licensing to upstream manufacturers may constitute appropriate public policy. Other commentators have even argued that refusing to license upstream component manufacturers may constitute an abuse of dominance and thus infringe the competition laws.[3]

The paper is structured in two parts. We start by showing that in a frictionless environment – understood as an environment where all information is public, firms are free to set prices for the goods they sell, and negotiation among firms jointly maximizes the benefits of the parties involved – the way in which a royalty rate is structured does not distort competition nor diminish social welfare. Thus, in this ideal setting the patent holder cannot use the royalty structure in an opportunistic way to affect market structure or extract additional rents

[1] Quanta Computer, Inc., et. al. v. LG Electronics, Inc., 553 U.S. ___, 2008 U.S. LEXIS 4702 (June 9, 2008). The court found that patent exhaustion does indeed apply to method patents and thus applied to LG Electronics' patents as embodied in Intel's components sold to Quanta Computers. For a sample of the debate that has continued after the Supreme Court's ruling in the case, see Jason McCammon, "The Validity of Conditional sales: Competing Views of Patent Exhaustion in Quanta Computer, Inc. v. LG Electronics, Inc, 128 S. Ct. 2109 (2008)", *Harvard Journal of Law & Public Policy* (2009); Sue Ann Mota, "The Doctrine of Patent Exhaustion: Not Exhausted by the Supreme Court in Quanta Computer v. LG Electronics in 2008", *SMU Science & Technology Law Review* (2008); Erin Austin, "Reconciling the Patent Exhaustion and Conditional Sale Doctrines in Light of Quanta Computer v. LG Electronics", *Cardozo Law Review* (2009); John W. Osborne, "Justice Breyer's Bicycle and the Ignored Elephant of Patent Exhaustion: An Avoidable Collision in Quanta v. LGE", *The John Marshall Review of Intellectual Property* (2007-08); and James W. Beard, "The Limits of Licensing: Quanta v. LGE and the New Doctrine of Simultaneous Exhaustion", *UCLA Journal of Law & Technology* (2008).

[2] Brian Love, "Patentee overcompensation and the entire market value rule", *Stanford Law Review*, (2007).

[3] Sven Völcker, A few questions regarding the European Commission's antitrust enforcement with respect to standards- Studienvereinigung Kartellrecht International Forum on EU Competition Law 2014.

from downstream competitors or consumers. In the second part of the paper, we discuss the impact of some of the frictions that are, in reality, typically present in technology markets. Even in this less-than-ideal environment, we show that it is still often in the interest of both the patent holder *and* society as a whole to split the total royalty burden among different parties in the production process.[4]

For the first part, we start by assuming a Pareto optimal bargaining process between upstream and downstream producers. That is, we assume that the negotiations taking place along the vertical supply chain exhaust all possible gains to both parties so that no money is left on the table. We find that different assignments of the royalty burden across the vertical supply chain have no impact on social welfare and that it is instead the aggregate royalty per unit sold in the downstream market that matters. We illustrate this general principle, which we denote as *royalty allocation neutrality*, first by means of example. We show that neutrality operates in typical vertical relations regardless of the market power the upstream manufacturer might have, the necessary investments for production, or the possibility that the upstream firm might also integrate a downstream production unit.

To provide intuition for this result, consider a situation where a patent holder charges a royalty to both an upstream and downstream producer. The upstream producer sells its component, which incorporates the patented technology, to the downstream producer, who then sells the end good to consumers. What happens if the patent holder shifts part of the royalty burden upstream, raising the upstream royalty but lowering the downstream one so as to leave the aggregate royalty unchanged? It turns out that the intermediate price the upstream manufacturer charges the downstream party will rise to accommodate the increase in costs that the upstream manufacturer incurs. In other words, the upstream manufacturer passes on the additional royalty payment, exactly offsetting the cost-savings enjoyed by the downstream producer from the lower royalty rate. This pass-through element is the fundamental insight of Ronald Coase in his famous theorem on the reallocation of costs to achieve an economically efficient outcome.[5] The reallocation argument only holds, though, if there are no transaction costs or roadblocks in the way of passing costs down the chain.

In many relevant instances, of course, roadblocks or transaction costs are important. Therefore, we devote the second part of the paper to discussing some of the likely implications of transaction costs on the optimal allocation of royalties. In particular, we discuss two such frictions: the existence of private information and the constraints firms may face in pricing the final good (for example, if the upstream producer cannot price discriminate among different downstream buyers).

To see how frictions can alter the analysis, consider the case of private information. Royalty payments are typically predicated on the amount of a good actually sold in the marketplace. The literature has proposed many explanations for the predominance of these contracts, mainly related to private information.[6] As opposed to the sale of a physical good where total sales can be estimated from the units of the input transferred, intellectual

---

[4] In both parts we are evaluating ex ante licensing negotiations, before potential licensees have begun to make any products that might implement the patented technology and thus before any irreversible investments have been made.

[5] See Ronald H Coase, "The Problem of Social Cost", *Journal of Law & Economics*, (1960).

[6] See Gerard Llobet and Jorge Padilla, "The Optimal Scope of the Royalty Base in Patent Licensing", Available at SSRN: http://ssrn.com/abstract=24172162014, and references therein.

property allows for an unlimited number of units. That means patent holders must be able to verify the quantities sold. It also means that licensees can have incentives to underreport sales in order to reduce their royalty payments. In this case, enforcing the patent contract is more complicated and the patent holder may strictly prefer to contract with multiple layers in the production chain, charging each link a partial royalty. This is because the lower individual rates reduce licensees' incentives to underreport royalty payments owed and the multiple licenses provide several check points for verifying quantities sold.

Overall, the lesson from our analysis is that in most cases the division of the royalty among the different firms in the production process has no impact on social welfare. Different allocations will be more or less desirable inasmuch as the total royalty they represent is higher or lower than what is socially desirable. Charging just one versus charging multiple parties is not the pivotal element for social welfare. In fact, in the presence of frictions, the way the total burden is split is likely to reflect the cheapest and most convenient way to implement licensing, which is bound to differ across firms, industries, and sectors of the economy. In some occasions charging multiple parties might be crucial to maintaining both economically justified rewards and efficient licensing. In others, social welfare might be maximized if patentees choose to license upstream producers or downstream producers only.

The remainder of the paper is organized as follows. Section 2 offers some examples of royalty allocation neutrality to clarify our ideas. Section 3 then presents the general principles underlying royalty allocation neutrality. Section 4 considers the implications of dropping the assumptions required to obtain royalty allocation neutrality. We conclude in Section 5.

2. Some Examples of Royalty Allocation Neutrality

In the introduction we claim that *royalty allocation neutrality* is a general property holding in markets where a patent holder can charge royalties to firms that operate in different stages of the production process. In this section, we discuss several examples that provide intuition on the mechanism at work. In the next section we discuss the general model and the assumptions behind the royalty allocation neutrality result.

To illustrate our analysis, we introduce the following generic market structure. Consider a market where three firms operate: a downstream producer *D* sells a final product to consumers according to a demand function $D(p)$; an upstream producer *U* sells a necessary input required by *D* to create the final product (*U* might also sell a product in the final market); and a patent holder *H* that can license its patent for a cost-reducing innovation *P* to both *D* and *U*. For every unit of the final good that *D* produces, it needs one unit of *U*'s intermediate good, for which *U* charges s. The total cost of producing the final good without the patented technology is c_0. If the final good is produced with the patented technology, the total marginal cost falls to $c < c_0$.[7] In exchange for a license

[7] Throughout the paper we will assume that c is sufficiently small so that even when the upstream producer charges its monopoly price, production with the alternative technology is unprofitable. In example 1, however, we show that the results do not hinge on this simplifying assumption.

to use P, the patent holder H demands a royalty r_U from the upstream input producer U and a royalty r_D from the downstream producer D.[8] The first example describes the simplest scenario.

Example 1 (Bilateral Negotiation): Suppose that the intermediate input price s is the result of negotiation between the upstream monopolist producer U and the downstream monopolist D.[9] Furthermore, this negotiation is carried out following Nash Bargaining, where U has bargaining power β ($0 \leq \beta \leq 1$) while D has the reciprocal power $1-\beta$. For simplicity, we postulate a linear demand function $D(p) = 1-p$, although the result applies generally. The timing is as follows. The patent holder H first chooses royalties r_U and r_D. The upstream and downstream firms then bargain over s and then the downstream producer chooses the final price p.

Starting from the last stage, D chooses the price according to its profit function, $\max_p (p-(c+s+r_D))D(p)$, which is simply the price net of costs multiplied by quantity demanded and which results in the profit maximizing monopolist price $p^M(s) = (1+c+s+r_D)/2$. With this price, the upstream producer obtains profits $\Pi_U(s) = (s-r_U)D(p^M(s))$, which again is simply the upstream price net of upstream costs multiplied by quantity demanded at the given downstream price. Notice that without loss of generality we have normalized the cost of the upstream producer to 0.

Working our way up the production chain, the equilibrium price for the upstream component s is then determined by maximizing the function $\max_s (\Pi_U(s))^{\beta}(\Pi_D(s))^{1-\beta}$, which yields the optimal intermediate price $s^* = (\beta(1-c-r_D) + (2-\beta)r_U)/2$. This optimal intermediate price is increasing in the bargaining power of the upstream producer in the relevant range (i.e., the higher β, the higher s^*). Note that to the extent that bargaining power is positive, $\beta > 0$, the optimal intermediate price will strictly exceed the upstream royalty rate ($s^* > r_U$), resulting in the familiar problem of double marginalization. Furthermore, s^* increases when the downstream royalty (r_D) decreases, or when the upstream royalty (r_U) increases.

Replacing $p^M(s)$ and using the demand function we obtain the optimal quantity sold as $q^* = (2-\beta)(1-c-r_D-r_U)/4$. As the equation shows, the optimal quantity is decreasing in the total royalty, $r_U + r_D$, which operates to increase the marginal cost of the downstream firm. Optimal quantity is also decreasing in the bargaining power of the upstream producer, β.

Profits for the patent holder H can be written as $\Pi_H = (r_U + r_D)q^*$, where q^* is, as observed before, a function of only the sum of r_D and r_U. That is, the distribution of the royalties between U and D is neutral from the patent holder's perspective, only their sum matters.[10] In fact, it is easy to verify that the optimal sum of royalties is characterized by

$$(1.0) \qquad r_U^* + r_D^* = \frac{1-c}{2}.$$

---

[8] We abstract from all other license terms, but in reality most license contracts contain a host of provisions of which royalty rates are but one.

[9] Since D is the only downstream producer, it is also a monopsonist buyer for U's good.

[10] As pointed out before, these results generalize to other demand functions; e.g. the isolelastic function.

Finally, notice that the previous computations assume that s^* is such that the downstream producer prefers not to buy the upstream product based on the alternative technology, competitively priced at $s_0 = c_0$. In particular, the downstream firm will buy the alternative input only if $s^* > c_0 - c - r_D$. If the intermediate price that results from the Nash bargaining does not satisfy this constraint, firms will optimally set $s = c_0 - c - r_D$. Substituting the final quantity sold we obtain $q^* = (2 - \beta)(1 - c_0)/4$. Hence, the patent holder will maximize its profits $\Pi_H = (r_U + r_D)q^*$ subject to $r_U \leq s$, where the constraint ensures that the upstream producer obtains non-negative profits. It is easy to see that the optimal combination of royalties is equal to the production cost savings, so that $r_U^* + r_D^* = c_0 - c$, which, again, only depends on the sum of royalties.[11] ■

Several things are worth pointing out from the previous example. In the negotiation process between the upstream and the downstream producer, the intermediate price *s* exhausts the entire surplus that both producers can achieve. In other words, the Nash Bargaining solution guarantees efficiency of the outcome. The solution divides surplus depending on how the bargaining power is allocated.

The bargaining power of each of the two parties is independent of the way that royalties are allocated. In this way we rule out, for example, the unrealistic possibility that the upstream producer is the one choosing the intermediate price *s* ($\beta = 1$) for some values of r_D and r_U, whereas the downstream producer chooses the price for other values ($\beta = 0$). Our setup nonetheless allows for the possibility, for instance, that the upstream producer is limited in its pricing choice by the presence of an alternative input technology.

Finally, it is important to note that, although the intermediate price *s* leads to an efficient allocation for the coalition between the upstream and the downstream producer, the price chosen in the downstream market does not necessarily accomplish the same goal. In fact, in the previous example, the price that maximizes profits where both the upstream and downstream firms are integrated corresponds to $p^M(s) = (1 + c + s + r_D)/2$. When firms are separate, however, the double-marginalization distortion mentioned above leads to a higher final price. This result would suggest that charging all of the royalty downstream could benefit the patent holder. As it turns out, however, this conjecture is not true, since the intermediate price *s* adjusts as the upstream royalty rate changes. So, an important lesson is that double-marginalization distortions are unrelated to the contract that the patent holder might be able to offer and derive solely from the relationship between *U* and *D*.

Note that the result regarding double marginalization does not hold if the patent holder could charge a two-part tariff, consisting of a royalty rate and a fixed fee. Even though the same royalty allocation neutrality result would hold, because the patent holder could extract its rents through a fixed fee that does not affect marginal quantity decisions, the license could undo the double-marginalization arising from the relationship between the upstream and the downstream producer by charging a *negative* royalty rate upstream. In that case the downstream monopoly price (which is lower than the price under double marginalization) would be attained, downstream quantities would be higher, the upstream producer would be compensated for its inability to charge

[11] If there are recurring costs to licensing, those would need to be included. Further, we assume here that the patent holder's initial investments in R&D are covered by the aggregate royalty rate.

a margin through the negative royalty rate, and the patent holder would be compensated through the corresponding fixed fees. That said, while theoretically possible, negative royalty rates seem unrealistic.

Example 2 (Raising Rival's Cost): Suppose that in the previous example the upstream producer, firm *U*, also produces a substitute good to the one sold by the downstream firm *D*. It is well-known that the upstream producer might raise the price of the input of the downstream competitor as a way of increasing the benefit *U* gains from the sale of its substitute product.[12]

To simplify the exposition and make the case for the raising-rival's-cost argument more obvious,[13] we assume that the upstream producer has all the bargaining power, $\beta = 1$, although the result would hold for a general β. Both downstream competitors face a symmetric downstream demand $D_i(p) = 1 - p_i - dp_j$ for $i, j = U, D$ and $i \neq j$, where the parameter $d \in (0,1)$ measures the degree of substitution between the two products.[14] The upstream producer sells the final good at a price p_U and incurs a marginal cost of production $c_U = c + r_U + r_D$. Note that the cost of the upstream producer includes both $r_U + r_D$ because this firm is vertically integrated.

We adapt the timing of the model in the natural way. First, the patent holder decides the royalties it charges to both upstream and downstream producers. Second, the upstream producer chooses the price for the intermediate input *s* and, finally, both manufacturing firms simultaneously choose end market prices, p_U and p_D.

We start from the final stage of the game. Both firms choose prices to maximize profits. In particular, profits for the upstream producer correspond to

$$(1.2)\quad \Pi_U(s, p_D) = \max_{p_D} (s - r_U) D_D(p_U, p_D) + (p_U - p_D) D_U(p_U, p_D).$$

Notice that by choosing the price downstream (p_D), the firm *U* also affects how much it receives from its downstream competitor through the units it sells in the final good market. Solving this expression we can observe that the equilibrium in this last stage leads to prices p_U^* and p_D^*, both of which increase as royalties (r_U, r_D), the intermediate price (*s*), the degree of substitution (*d*) or the cost of production (*c*) increase.

If we replace the prices in equation (1.2) above with their formulae and then maximize with respect to the intermediate good price *s*, we are able to obtain the optimal intermediate price $s^*(r_U, r_D, c, d)$, which decreases as either c or r_D increases and which increases as r_U increases, if substitution d is sufficiently large. Furthermore, it can be shown that $s^* + r_D$ depends only on aggregate royalties, $r_U + r_D$, not on either of the royalties individually.

Using the previous result, we obtain that the marginal cost of the downstream producer, $s^* + r_D + c$, is only a function of the sum of the royalties, in the same way that the production of a unit for the upstream integrated firm entails a royalty payment of $r_U + r_D$. It is immediate, therefore, that the equilibrium quantities q_U^* and q_D^*

---

[12] See, for example, Janusz A. Ordover, Garth Saloner and Steven C. Salop, "Equilibrium Vertical Foreclosure", *American Economic Review*, (1990).

[13] This term was coined by Salop and Scheffman (1983). When this increase in costs prices the otherwise efficient competitor product out of the market it is usually referred as *vertical price squeeze*. Steve C. Salop, S.C. and David T. Scheffman. "Raising Rivals' Costs", *American Economic Review* (1983).

[14] All the results carry over to the case where products are complements, represented by $d < 0$.

are a function only of the sum of the royalties. Since profits for the patent holder arise in this case from total royalties multiplied by total quantities, $\Pi_H = \max_{r_U, r_D} (r_U + r_D)(q_U^* + q_D^*)$, it must be the case that only the sum of the royalties matters. In particular, it can be shown that profits are maximized if the sum of royalties is equal to a particular combination of costs and substitution, given by $r_U^* + r_D^* = (1 - c)(1 - d)/2(1 - d)$. Notice that the sum of these royalties falls as the marginal cost of production c falls and as the degree of product differentiation falls (or in other words, the sum of royalties rises as the degree of substitution d increases). Of course, when products are independent $d = 0$ and this expression corresponds to the one given by (1.0). ■

The previous example shows why the principle of royalty allocation neutrality holds even in markets where upstream manufacturers are vertically integrated and there are a number of downstream competitors. Even when the patent holder charges a different combination of royalties up and downstream, the upstream manufacturer responds by changing the wholesale price s so as to leave the total input cost for the final product, $s^* + r_D$, unchanged. It is in this sense that we say the aggregate royalty is the key variable, not the individual rates.

Example 3 (Double-Sided Moral Hazard): Consider next a model where effort plays a role in product success. Suppose that in the first example, before the upstream and downstream producers bargain over the price s at which they will exchange the input in the final production process, both firms need to make an investment. The cost of this investment is e_i (for i = U, D) and it leads to a probability of success, $P(e_U)\,P(e_D)$, where P is increasing and concave in its argument.[15]

Net profits for the upstream and downstream producer are equal to the probability of success multiplied by the normal profit (which is a function of the royalty rates paid) minus the effort investment. This can be written as, respectively for the upstream and the downstream firm, $\Pi_i(r_U + r_D) = \max_{e_i} P(e_U)\,P(e_D)\,\pi_i(r_U + r_D) - e_i$, where the profit functions $\pi_U(r_U + r_D)$ and $\pi_D(r_U + r_D)$ are the profits obtained when the investment is successful. Recall that we demonstrated above that these two equations depend only on the sum of the royalties. It is immediate, then, that the efforts exerted by the upstream and downstream producers are independent of the way the royalties are allocated among the two and again only the sum matters. Of course, the same result would arise for more general forms of complementarity between the effort of the upstream and the downstream producers or even in the case in which this complementarity does not exist.

Finally, notice that this setup also includes the case where only the investment of one of the parties is required, as it would correspond to $e_i^* = 0$ for the other one. ■

This last example illustrates why the neutrality principle still holds in the presence of a moral hazard. As long as the Nash Bargaining process works to maximize the two manufacturers' joint profits, there is no role for the patent holder H to manage effort levels by manipulating the royalty rates. These few examples lead us to a general principle of royalty allocation neutrality, which we explain next.

[15] Lafontaine (1992) shows that double-sided moral hazard characterizes, for example, the relations between franchisees and franchisors. Francine Lafontaine, "Agency Theory and Franchising: Some Empirical Results", *RAND Journal of Economics*, (1992).

3. The General Principle of Royalty Allocation Neutrality

The previous examples, although quite different in nature, share two common features that turn out to be crucial for royalty allocation neutrality. The first is the fact that different allocations of individual royalties do not lead to changes in the overall profits attainable by a coalition between the upstream and downstream producer. In other words, the same sum of profits can be achieved by appropriately changing the intermediate good's price s for different divisions of the same total royalty burden.

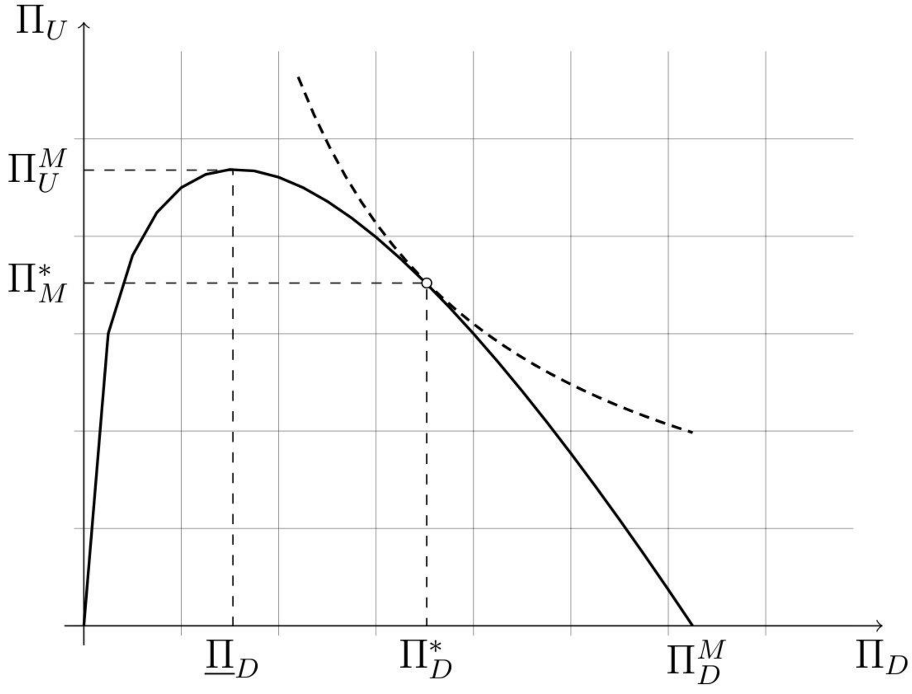
The second feature has to do with the way the total surplus is allocated among the parties. In the previous three examples, the allocation of the surplus is efficient. Using the *Coase Theorem*, and its famous insight on the reallocation of costs, it is immediate that the total final quantity should be independent of who is paying the cost of this production.[16] Most important for our purposes is the fact that the rule governing the allocation of the surplus is independent of how the royalties are allocated. As pointed out before, however, if the upstream producer had price setting power for some values of r_D and r_U but not for others, the results would certainly vary. We find this scenario unlikely though.

Figure 1 illustrates these two features. The solid line corresponds to all the possible allocations of profits among the upstream and downstream producers. These allocations are spawned by the different values of s. Profits for the upstream producer range from Π_U^M, which are the profits obtained when it has all the bargaining power, to 0, which hold when $s = r_U$ or else when s is so high that the quantity sold is 0. Notice that the first case corresponds to the maximum profits that the downstream producer can achieve (in the figure labeled as Π_D^M). In the other case both profits are 0. Efficient negotiation rules out all values of s higher than the monopoly one, which generate outcomes in the region where $\Pi_D^* > \underline{\Pi}_D$.

It is important to notice that for a given level of $r_U + r_D$ this set is unchanged regardless of the way the individual royalties are allocated. In other words, if this set arises for a certain combination (r_U, r_D), it will also arise for the combination (r_U', r_D') as long as the sum is the same (i.e., $r_U' = r_U + \varepsilon$ and $r_D' = r_D - \varepsilon$) and the wholesale price s is simply replaced by $s' = s + \varepsilon$.

[16] Notice that we are making use of the so-called "strong" version of the Coase Theorem that states not only that the outcome is efficient but also invariant to the assignment of property rights. Here, the theorem applies to the relationship between the upstream and the downstream firm. Because the negotiation process leads to an efficient outcome that divides the total surplus according to the bargaining power of each of the parties, the final allocation only depends on the total royalty payment and not on how the patent holder structures these payments among the two firms. The price of the input adjusts to any change in royalty structure, leading to an invariant allocation (in this case, the quantity being produced).

Figure 1: Division of profits for a given royalty sum $r_U + r_D$



The dashed curve in Figure 1 shows the objective function that the Nash Bargaining solution maximizes. The point at which both curves are tangent determines the outcome. A higher value of β, meaning more bargaining power for the upstream producer U, leads to a higher value on the solid line, getting closer to the upstream monopoly allocation $(\underline{\Pi}_D, \Pi_U^M)$, achieved when β=1. In other words, when the upstream firm behaves as a monopolist it charges an intermediate price sufficiently high so as to maximize its profits. Of course, if s is too large, profits for both parties decrease. More bargaining power for the downstream producer (a lower β) generates allocations that tend towards downstream monopoly $(0\ \Pi_D^M)$, , achieved only when $\beta=0$. Clearly, for the allocation of royalties to be neutral we require that bargaining power β does not depend on either royalty rate (r_U or r_D), or alternatively that β depends only on their sum, but this independence strikes us as a natural assumption. The previous intuitions allow us to state the following general result (proven in appendix A).

Proposition 1 (Royalty Allocation Neutrality): If the final price of each good is increasing in its marginal costs and the negotiation between the upstream and downstream producers has the following properties:

1. the outcome is Pareto optimal for the coalition, and
2. expansions of feasible combinations of profits cannot make any of the parties worse off,

then, the final allocation depends only on the sum of the royalty charged for the upstream and the downstream producer of each good.

Notice that our previous examples naturally satisfy these conditions. The Nash bargaining solution that is assumed (cases where the upstream firm acts as a monopolist or the downstream firm as a monopsonist are particular examples) always leads to efficient outcomes. Furthermore, the price of the intermediate good is such that each firm's profits are increasing in the total surplus divided between the upstream and the downstream firms according to the bargaining weight β.

The proof developed in the appendix is based on the fact that under the conditions stated in the text of this proposition, the upstream producer optimally adapts the intermediate price, as shown in example 2, to exactly offset the impact that changes in the distribution of the total royalty have on the marginal cost perceived by the downstream producer. In other words, the upstream producer optimally chooses s^* so that the sum $s^* + r_D$ depends only on the sum of the royalties. Thus, the final price and the total number of units sold will change only if the sum of the royalties change.

4. Deviations from Royalty Allocation Neutrality

As we explained in the previous section, royalty allocation neutrality hinges on the flexibility that the upstream and downstream producers have to adjust the wholesale price s in order to compensate for different royalty allocations by the patent holder. Implicit in this argument is the idea that negotiation is frictionless and its outcome cannot be improved upon. But oftentimes it is not realistic to assume that negotiation will lead to a Pareto optimal outcome; the quest for individual profits might destroy overall value. Similarly, Pareto efficiency might break down due to technological or institutional constraints that may limit pricing flexibility, for example, for the upstream producer vis-à-vis the downstream manufacturer. In those cases, s cannot be adjusted at will to maintain joint profit maximization. In this section we elaborate on these two reasons for royalty allocation neutrality to fail, and we illustrate them with variations of Example 1.

Regarding the first case, we focus on information asymmetries that are likely to appear and that might affect contract enforcement. 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 – 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.

Oftentimes the relevant quantity sold is not easily verifiable. Centralized exchanges are not very common and when they do exist they only channel a limited proportion of the total units sold. The bulk of transactions are carried out using private long-term contracts. But even with established long-term contracts, for many technology products the sale of intermediate components is difficult to monitor. In fact, this is an important concern for many firms and has triggered the development of tracking systems for individual units (or boxes) being sold for licensing purposes. For example, one firm instituted a license whereby a box can only be shipped if proof of the license is displayed on the outside.[17]

Even where monitoring sales is relatively easier, it can be difficult to identify the exact goods covered by a license. Manufacturers often sell a myriad of similar yet different products, so determining which do and which do not incorporate the intellectual property licensed can be difficult and expensive. This is especially problematic for complex high technology products, such as goods that incorporate multiple semiconductor chips and thus tend to rely on hundreds, if not thousands, of patents.

As a result of these complications, when the relevant quantities are not directly observable the patent holder may choose to monitor one or several stages of the production process to improve information collection. For example, if monitoring is costly, it may make sense to concentrate effort in the stage where this cost is lowest. The cost of monitoring the level of production might be different depending on the number of players in the market or the closeness 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 is likely to take place in this last stage.

More interestingly, the patent holder might often want to charge royalties at the different stages of the production chain if splitting fees in this fashion allows the firm to obtain additional estimates of the quantity sold that complement the direct observation gathered through monitoring. Alternatively, lower royalty rates to each stage might increase the odds of obtaining accurate information (i.e., reduce the incentives of licensees to under-report). Timing might play a role as well if negotiations with a downstream producer, say, can take place at a later date, when the downstream firm knows its needs better, whereas at the time of negotiations with the upstream producer the downstream needs would need to be forecasted.

Furthermore, the patent holder may prefer a spread allocation of the royalty burden due to the differences in the cost of enforcing a licensing contract. 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 for which the product incorporating the patent is highly profitable. This threat alone might often be enough to allow the proper enforcement of the contract. As a result, if enforcement is an important concern, the patent holder may want to shift the royalty burden towards those stages in which licensee competition is weaker and the price margin is higher. In technology markets, these high-margin stages are often the ones closest to the final consumer, where product differentiation makes competition less fierce. The next example outlines many of the forces that we have discussed so far.

Example 4 (Asymmetric Information): Consider the model presented in example 1, and for simplicity assume that $\beta = 1$ so that the upstream producer has all the bargaining power. Assume that the patent holder does not

---

[17] Such a per batch licensing system has been introduced by Phillips (who dubbed 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 that signals to the traders and retailers that the merchandise is licensed.
See Philips Intellectual Property & Standards, Licensing at
https://www.ip.philips.com/services/?module=IpsLicenseProgram&command=View&id=20&part=7).

observe the final quantity sold, which is only observed by the upstream and downstream producers. The patent holder knows, however, that the demand for this product will be high (represented as $D(p) = 1 - p$) with probability ν and low (represented as $D(p) = 0$) with probability $1 - \nu$.

High demand corresponds to our original Example 1, thus the upstream and downstream firms will operate under the same strategy as they did there. Denote their profits, gross of royalty payments, as π_U and π_D. When demand is zero, the prices chosen will be irrelevant and profits will be 0. The patent holder, however, absent a monitoring mechanism cannot distinguish between these two outcomes and thus will never obtain any royalty payments, since both firms will claim that the low realization of the demand occurred in order to avoid paying royalties. Profits for the patent holder are thus 0.

Suppose now that the patent holder has access to a monitoring technology, such that spending amounts k_U and k_D allows him to verify through the upstream or the downstream stage, respectively, which realization of the demand occurred. Clearly, in this case, the patent holder will maximize profits by monitoring only the stage of the process where this cost is lowest.

Alternatively, the patent holder might be able to impose a cost on firms misrepresenting their sales figures. For illustration purposes we assume that this cost is fixed and equal to $F > 0$. The patent holder can use this threat to elicit the private information that licensee firms possess. (See Appendix B for an application of our model framework to private information.)

In particular, consider the following contract that the patent holder can establish. A royalty is charged in both stages of the process. The patent holder requires each firm to declare the number of units being sold. If both firms make the same assessment, the upstream and downstream firm pay a per unit royalty r_U and r_D. If firms differ in their assessments, the firm that has declared a demand of 0 is fined an amount F. The other firm pays a royalty for the units it declares to have sold. The payoff matrix when the high demand realization has occurred reads as shown in Table 1.

Table 1. Payoff Matrix

		D: High	D: Low
U	High	$\pi_U - r_U q, \pi_D - r_D q$	$\pi_U - r_U q, \pi_D - F$
	Low	$\pi_U - F, \pi_D - r_D q$	π_U, π_D

Notice that with the existence of fines, declaring that demand was high becomes a Nash Equilibrium as long as F is sufficiently high. In cases where charging the total royalty in one of the stages does not satisfy these constraints, the allocation that splits it in the different stages may prevent firms from misrepresenting their sales.[18] ■

The second source of friction that might break royalty allocation neutrality is related to technological or institutional limits in the way that the intermediate price may adjust to changes in the royalties. In the limiting case, if U had no control at all such that s were exogenously determined, and no pass-through were possible, different ways to split the total royalty would lead to different final allocations. More interestingly, consider the case where a single upstream producer sells to several downstream firms. In this case, royalty allocation neutrality 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 it only charges a royalty upstream and the upstream producer modifies the price of the intermediate product appropriately. However, it is often the case that the upstream producer is serving several downstream producers for which the same input has a different added value. If, due to arbitrage or antitrust considerations, the upstream firm cannot charge different prices downstream, in contrast to what a patent holder might be able to accomplish using individual licensing negotiations, royalty allocation neutrality will fail. Furthermore, the upstream producer will tend to raise its intermediate price in a way that makes the double-marginalization problem more severe, reducing as a result both profits and consumer welfare.

Example 5 (Arbitrage): Go back again to example 1 in a much simpler setup, where the upstream producer has all of the bargaining power, $\beta = 1$, and the licensed technology brings production costs down to zero, $c = 0$. Assume that the upstream firm serves two downstream producers. Each downstream producer sells to one consumer willing to buy one unit. The consumer in market 1 has a valuation θ_1, while the consumer in market 2 has a lower valuation, $\theta_2 < \theta_1$. It is immediate that the optimal contract in this case will call for the sum of the royalties to equal each valuation in turn, with intermediate prices s adjusting accordingly: $r_{U,1} + r_{D,1} = \theta_1$, $r_{U,2} + r_{D,2} = \theta_2$, $s_1 = r_{U,1}$ and $s_2 = r_{U,2}$.

Assume now that the upstream producer, due for example to arbitrage, cannot price discriminate between the two downstream firms and is forced to sell in both markets at the same intermediate price, $s_1 = s_2$. The optimal allocation can still be attained if all the royalty burden is shifted downstream so that $r_{U,1} = r_{U,2} = 0$ and thus $s_1 = s_2 = 0$. If this possibility is precluded, however, the upstream producer will need to choose between an intermediate value of s that will reduce profits from market 1, or if the difference between the two downstream valuations θ_1 and θ_2 is sufficiently high, the upstream producer can decide to sell only in market 1. As a result, the patent holder will obtain lower profits and social welfare will surely decrease since market 2 will be left underserved or unattended altogether. ■

As the previous example points out, in practice, this sort of friction is likely to arise when there is important heterogeneity in the uses (and the corresponding valuation) of the intermediate good produced by the upstream firm. This is likely to be the case for goods that have a general span of uses or markets for highly differentiated

[18] Of course, even when this penalty exists there is a Nash Equilibrium where both firms declare that demand was low regardless of a true high realization.

final products. The prediction of the model is that in this case the patent holder will tend to shift (at least some of) the royalty burden downstream. In this situation, different contracts can be written for firms with different needs.

5. Conclusions

Motivated by the arguments raised during the *Quanta* case, we have attempted to present a comprehensive analysis of the economics of patent licensing in a vertical relations environment. In an ideal setting, with no transaction costs or asymmetries of information, we find that the division of royalties among the different firms in the production process has no impact on social welfare. Different allocations will be more or less desirable inasmuch as the total royalty they represent is higher or lower than what is socially desirable. Charging just one versus charging multiple parties is not the pivotal element for social welfare.

But the idealized world presented in our benchmark case is likely to differ from the market reality. For instance, technology markets may be characterized by private information or wholesale pricing constraints. Even here, though, there is no justification for placing restrictions on the ability of patent holders to split fees among multiple production layers. In fact, in the face of transaction costs and frictions, a strict interpretation of first sale patent exhaustion is likely to generate welfare losses in the economically justified reward and efficiency dimensions of licensing discussed in the introduction. The reward from innovation decreases because the patent holder will likely obtain lower profits and efficiency will be reduced if the restrictions reduce the diffusion of innovations in downstream markets.

We have identified several environments where the costs of a strict application of the first sale doctrine are likely to be relevant. The main one is related to informational asymmetries between the patent holder and the upstream and downstream producers. These asymmetries are likely to be particularly relevant in the case of technologies that have a range of applications, that can be embedded in a variety of products or that are used in combination with other products. Similarly, the reward that patent holders receive from their innovations is likely to be affected by licensing restrictions when the need to price discriminate in the downstream market is more important. In that case, a patent holder might do better in terms of screening different applications of the patented technology, in ways that upstream producers might be unable to replicate.

We have also found that when volume information is private or when the upstream producer cannot price discriminate between downstream producers, it may be in the interest of both the patent holder and consumers to allocate all the royalty burden to the downstream producer. Our findings fail to support the position of those that claim that a policy that restricts licensing to upstream manufacturers in cases involving products that integrated many components may constitute appropriate public policy. Our results also show that refusing to license to upstream component manufacturers may be welfare enhancing and, in that case, would not infringe competition laws.

Our analysis also applies to other forms of intellectual property. For example, the issues addressed in this paper regarding the division of the total patent royalty are closely related to those that arise in the literature on the European doctrine known as droit de suite (DDS). The DDS rule – which began in France in the 1920s and over time was adopted in many European countries – guarantees a royalty to the creator of a work of art over the price obtained in its future resale. Some authors, such as Ginsburg (2005),[19] have raised a 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 where artists do not receive such a royalty. In contrast, applying our royalty neutrality analysis to DDS implies that the prices initially paid to an artist would adjust to reflect expected DDS payments made in the future. Empirical studies of DDS rules are consistent with our conclusions. In particular, the evidence presented in papers such as Graddy and Banterghansa (2011) show that the impact on future prices for artwork covered by DDS is negligible, as are any trading volume effects.[20] If anything, the quantitative studies indicate that trading volume increases more in DDS-friendly countries.

[19] Victor Ginsburgh, "The Economic Consequences of Droit de Suite in the European Union", *Economic Analysis & Policy* (2005).
[20] Kathryn Graddy and Chanont Banternghansa, "The Impact of the Droit de Suite in the UK: An Empirical Analysis", The Journal of Cultural Economics, (2011), vol. 35 (2).

Appendix A. General Royalty Allocation Neutrality Argument

Proof of Proposition 1. Consider a market where firms *i=1,..,I* each produce one downstream product and *j=1,..,J* firms each produce one upstream component. A patent holder owns intellectual property necessary for the production of the good. Assume that the marginal costs of production are c_U^j and c_D^i for the upstream and the downstream good producer, respectively. The patent holder charges a royalty r_U^j and r_D^i to upstream and downstream producers. Finally, upstream producer *j* charges to final good producer *i* a price $s_{i,j}$.

Denote as $\left\{ r_U^{j*} \right\}_{j=1}^{J}, \left\{ r_D^{i*} \right\}_{i=1}^{I}$ a combination of royalties that maximizes profits for the patent holder. Denote as $q_{i,j}^{*}$ the quantity that downstream producer *i* buys from upstream producer *j* and $s_{i,j}^{*}$ its price. Profits for the patent holder can, therefore, be written as

$$\Pi_H = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(r_U^{j*} + r_D^{i*} \right) q_{i,j}^{*}$$

where

$$\sum_{j=1}^{J} q_{i,j}^{*} = D\left(p_D^{i*} \right) .$$

Royalty allocation neutrality means that any other combination $\left\{ r_U^{j\,'} \right\}_{j=1}^{J}, \left\{ r_D^{i\,'} \right\}_{i=1}^{I}$ with $r_U^{j\,'} \geq 0$ and $r_D^{i\,'} \geq 0$, and $r_U^{j\,'} + r_D^{i\,'} = r_U^{j*} + r_D^{i*}$ for all *i, j* leads to the same profits.

Towards a contradiction, assume that this is not the case and that the contract $\left\{ r_U^{j\,'} \right\}_{j=1}^{J}, \left\{ r_D^{i\,'} \right\}_{i=1}^{I}$ with upstream price $s_{i,j}^{'}$ and final sales sales $q_{i,j}^{'}$ produces strictly lower profits. Then, it must be the case that there exists at least a final good, say good *i*, for which total sales are lower under the alternative contract. Lower sales imply a higher final price for product *i*. Since profits are increasing in the cost of the product, there is at least one upstream producer, say firm *j*, for which

$$r_D^{i\,'} + s_{i,j}^{'} > r_D^{i*} + s_{i,j}^{*}.$$

That is, the downstream producer is facing a higher total licensing cost of buying from the original upstream firm, and it might (or might not) switch to another provider.

Under the alternative contract, the total profits that firm *i* and *j* split can be written as

$$p_D^{i'} - \quad c_U^{j} + c_D^{i} + r_i^{D'} + r_U^{j'} \quad q_{i,j}^{'},$$

where, by hypothesis, $q_{i,j}^{'} < q_{i,j}^{*}$. It is then, immediate, that the intermediate price $s_{i,j}^{'}$ does not maximize profits for the coalition between firm *i* and *j*, since by choosing

$$\hat{s} \equiv r_D^{i*} - r_D^{i'} + s_{i,j}^{*}$$

the total profits to be divided would have increased. Given that we have assumed that the profits of each of the firms cannot decrease when the feasible set expands, the original intermediate price $s_{i,j}^{'}$ could not have been Pareto optimal at this stage, leading to a contradiction.

Appendix B. Optimal royalties under private information

Proposition 2. Consider a private information game where the patent holder chooses in the first stage r_U and r_D, and where the upstream and downstream licensees do not observe the royalty imposed on the other firm. In the second stage the upstream producer chooses an intermediate price *s*. The downstream firm faces a demand function *D(p)=1-p*, as in example 1. Then, the aggregate royalty that maximizes the patent holder's profits is identical to the one obtained in the example 1.

Proof. Consider the case in example 1 with $\beta=1$. We start with the last stage of the game. Given r_D and *s* the downstream producer chooses a price for the final $p^* = \frac{1+c+r_D+s}{2}$. The quantity produced is $q^* = 1 - p^* = \frac{1-c-r_D-s}{2}$.

In the second stage, after r_U and r_D have been chosen, the upstream producer has beliefs ρ regarding the (unserved) r_D. Given those beliefs, the chosen *s* solves

$$\max_s (s - r_U)\left(\frac{1-c-\rho-s}{2}\right)$$

resulting in $s^*(\rho) = \frac{1-c-\rho+r_U}{2}$.

In the first stage, the patent holder maximizes

$$\max_{r_U, r_D} (r_U + r_D)q^* = \max_{r_U, r_D} (r_U + r_D)\left(\frac{1-c-2r_D+\rho-r_U}{4}\right).$$

It is easy to observe that for a given r_U+r_D profits are increasing if r_U is increased and r_D is decreased. Therefore, profits are maximized if is chosen at its minimum. That is, $r_D=0$. Since in equilibrium $\rho=r_D$ the profit function becomes

$$\max_{r_U} r_U\left(\frac{1-c-r_U}{4}\right)$$

which results in $r^*_U = \frac{1-c}{2}$ as stated.