Cumulative Innovation, Sampling and the Hold-Up Problem

Rufus Pollock

Cambridge University

24. January 2006

CUMULATIVE INNOVATION, SAMPLING AND THE HOLD-UP PROBLEM

RUFUS POLLOCK

CAMBRIDGE UNIVERISTY

V1.3.3 2007-08-10

Abstract. With cumulative innovation and imperfect information about the value of innovations, intellectual property rights can result in hold-up and therefore it may be better not to have them. Extending the basic cumulative innovation model to include 'sampling' by second-stage firms, we find that the lower the cost of sampling, or the larger the differential between high and low value second-stage innovations, the more likely it is that a regime without intellectual property rights will be preferable. Thus, technological change which reduces the cost of encountering and trialling new 'ideas' implies a reduction in the socially optimal level of rights such as patent and copyright.

Keywords: Cumulative Innovation, Hold-Up, Sampling, Intellectual Property

JEL codes: K3, L5, O3

Faculty of Economics, Cambridge University, Sidgwick Avenue, Cambridge, CB3 9DD. Phone: +44 (0)1223 690 423. Email: rp240@cam.ac.uk and rufus@rufuspollock.org. I thank my advisors, Rupert Gatti and David Newbery, for guidance and many helpful discussions. I also wish to thank Corinne Langinier and participants at seminars in Cambridge and Toulouse as well as those at the 2006 DRUID summer conference and the 2007 International Industrial Organization Conference for valuable comments and suggestions.

1. Introduction

> [The] 90-minute documentary [Wanderlust] ... was also a window into the frustrations of making a clip-intensive film dependent on copyright clearance, which has become hugely expensive in the past decade. Initial quotations for the necessary sequences came to more than $450,000, which would have raised by half the cost of the IFC film. ... "Paramount wanted $20,000 for 119 seconds of Paper Moon", Ms. Sams said. "The studios are so afraid of exploitation that they set boundaries no one will cross. Even after the prices were cut, we were $150,000 in the hole."[1]

Cumulative innovation and creativity, whereby new work build upon old, is a pervasive phenomenon. However, it was not until recently that it received significant attention in the literature. The seminal paper in this regard is that of Green and Scotchmer (1995). They introduced a two-stage innovation model in which the second innovation is enabled by, or builds upon, the first. Their paper primarily concerns itself with how rents are divided between innovators at the two stages, in particular with the extent to which the first innovator is (under-)compensated for her contribution (the option value) to the second innovation. They investigate how different policy levers related to intellectual property rights, in particular breadth[2], could be used to affect the bargaining (or its absence) between different innovators and hence the resulting payoffs.

A central feature of their model, as well as subsequent work that extended it (such as Scotchmer (1996)), was an assumption that knowledge of costs and returns, whether deterministic or stochastic, was shared equally by innovators at different stages (i.e. was common knowledge). With common knowledge all mutually beneficial transactions are concluded, using ex ante licenses where necessary to avoid the possibility of hold-up of second-stage innovators.

This assumption, however, is problematic. If all innovators share the same information why do we need different innovators at first and second stages and why concern ourselves

[1]The New York Times, May 28, 2006 *No Free Samples for Documentaries: Seeking Film Clips With the Fair-Use Doctrine.*

[2]A monopoly right (intellectual property right) such as a patent or a copyright confers the right to exclude not simply direct copies but also products that are sufficiently similar. The term lagging/leading breath are often used to denote the space of inferior/superior (respectively) products that are excluded by the patent/copyright (i.e. taken as infringing the monopoly).

with licenses and bargaining if a single innovator could just as easily do it all? The obvious answer is that this assumption is wrong, something suggested by a cursory observation of reality: many different firms engage in innovation precisely because they have specialized skills and knowledge that make it effective for them rather than another firm to engage in a given area.[3] Thus, in this paper we investigate cumulative innovation under asymmetric information, for example, where a first-stage innovator only has a probabilistic prior over the second-stage innovator's cost/values but the second-stage innovator knows them precisely[4].

Our paper takes as a starting point a 'basic' model very similar to that presented by Bessen (2004).[5] Second-stage firms are of two types (high and low value) with the type unobserved by first-stage innovators. With (strong) IP first-stage firms may require second-stage innovators to pay a royalty while with (weak) IP second-stage firms may produce without having to license from first-stage firms. As first-stage firms do not know the type of given second-stage innovator with (strong) IP there may be 'licensing failure' (that is the royalty may be set above the level that a second-stage firm is willing to pay).[6] Thus, there is a trade-off: with IP more first-stage innovation takes place due to the extra royalty income received by first-stage firms but some second-stage innovation may be lost as a result of 'licensing failure' due to high royalty rates.

Such a trade-off is already familiar in the literature and our main reason for presenting it is to provide a benchmark and basis for the more complex 'sampling' model presented in the second section. The 'sampling' model extends the first by introducing the idea of

[3]See e.g. Eisenberg and Heller (1998); Hall and Ziedonis (2001); Cockburn (2005).

[4]Of course, for consistency, the collective distribution of the values/costs of all second-stage innovators should correspond to the prior of the first innovator.

[5]We differ from Bessen slightly in that his focus is primarily on whether ex ante or ex post licensing occurs. Central to this analysis is his introduction of ex post royalty shares which are the royalty shares that take place in the absence of licensing. These are determined exogenously – perhaps as a policy variable or determined by invent-around costs and other factors – and Bessen shows that the socially optimal ex post royalty share is less than that obtained in ex ante bargaining (and so all licensing should occur ex post). By contrast, in our model we do not have the concept of an ex post royalty share: either a second-stage innnovator obtains a license or she does not (and so then cannot produce).

[6]We note that Bessen uses the term 'holdup' to denote what we term 'licensing failure'. Since he is considering ex-ante licensing his use of the term 'hold-up' differs somewhat from the traditional usage as there are no sunk relationship-specific investments (a binding contract is possible ex-ante). Rather the 'hold-up' is simply that, just like a monopolist facing heterogeneous consumers, a first-stage innovator is facing a set of second-stage innovators with private and heterogeneous values and so may set a profit-maximizing royalty rate that excludes some second-stage innovators from licensing. Since, the 'sampling' case we discuss below resembles more closely a traditional 'hold-up' situation we prefer to reserve that term for use there and to to use 'licensing failure' for the situation described here.

sampling, that is that second-stage firms engage in some form of (costly) effort prior to the point that any kind of royalty setting and licensing takes place. Imagine, for example, that a second-stage innovator must search out, trial and experiment with, (many) first-stage innovations prior to deciding which first-stage innovation they can, or want to, use (and therefore license). Furthermore, the more products they sample, the more likely it is a second-stage firm comes up with a good idea of its own (or a good match between its idea and existing ideas or products on which it can build) – which is modelled, in this case, by the firm's innovation being of high, rather than of low, net value (net, that is, of costs).

Classic real-world examples of such a situation can be found in the software and music industries. In software a new application will likely combine many ideas (and even code) from previous products. But ideas can only come from applications that one has encountered.[7] In music, particularly modern music, re-use either explicit or implicit is ubiquitous. For example, in dance and hip-hop, 'sampling', whereby a small section of a previous work is directly copied and then repeated or reworked in some manner, is the very basis of the genre. More generally all composers whether classical or modern use previous musical, ideas, motifs, and melodies as parts of new works[8].

Sampling benefits a firm by increasing the probability of having a high value innovation but it is costly. As it takes place prior to any kind of royalty negotiation it may lead to hold-up: the hold-up of the sampling effort. As a result the presence of IP rights that require second-stage innovators to license may now have another cost in addition to that from traditional 'licensing failure': fearing high royalty rates second-stage firms will reduce the level of sampling they do and thereby reduce the average quality of second-stage innovations. Because this effect operates across all second-stage innovators its consequences for welfare may be substantially greater than the traditional 'licensing failure' problem (which only affects low value second-stage innovators).

Turning to the comparative statics, we find that, in general, the lower the sampling costs or the larger the differential between high and low value second-stage innovations, the more likely it is that a regime *without* intellectual property rights will be preferable. Thus, in

[7]We should distinguish here between reuse of ideas and reuse of code. With software copyright but no software patents one would (within limits) be free to reuse the ideas found in existing applications. However, reuse of the code itself would require that the software be open-source.

[8]See e.g. Malcolm Gladwell, The New Yorker, 2004-11-22, *Something Borrowed: Should a charge of plagiarism ruin your life?*, also http://www.low-life.fsnet.co.uk/copyright/part3.htm#copyrightinfringement for information about sampling in dance and hip-hop music.

the context of this model, technological change which reduces the cost of encountering and trialling new 'ideas' should imply a reduction in the socially optimal level of intellectual property rights such as patents and copyright.

This approach therefore adds another dimension to the question of how profit is divided between innovators at different stages. Seen in this light, it also has direct analogies with existing results related to the question of whether second-stage innovations should be infringing (I) or non-infringing (NI). For example, Denicolo (2000), who extends Green and Scotchmer's model with patent races at each stage, finds that in some circumstances it will be better to make second-stage innovations non-infringing (in this model one trades off faster second-stage innovation with non-infringement against faster first-stage innovation when there is infringement).

It also has a close connection to the recent work of Bessen and Maskin (2006). Similar to this paper they investigate the welfare impact of 'licensing failure' due to asymmetric information in a model of cumulative innovation. Similar to us they show that, with cumulative innovation, in contrast to what occurs in a 'one-shot' model, IP may, in some circumstances, reduce rather than increase innovation (and social welfare). However their focus is rather different from ours (complementarities in research rather than sampling) and their results arise for different reasons. Specifically, in their model there are multiple stages with (the same) two firms at each stage. Each may choose to participate or not in researching the current innovation and the next innovation stage is reached if, and only if, research at the current stage is successful, with success an increasing function of the number of participating firms. As a result their is an 'externality' from participation in a given stage: though the value of success at the current stage accrues only to the winning firm by enabling subsequent stages (some of which may be won by the other firm) success also increases the other firms expected revenue. As a result, when one firm is excluded from subsequent stages due to 'licensing failure' under an IP regime the effect on welfare can be far more severe than in the one-stage case.

Finally, we should point out that our results are of relevance to a variety of recent policy debates. For example, in December 2006 the Gowers Review of Intellectual Property which had been setup by the UK government to examine the UK's current IP regime, provided, as one its recommendations (no. 11), that "Directive 2001/29/EC [the EU Copyright

'InfoSoc' Directive] be amended to allow for an exception for creative, transformative or derivative works, within the parameters of the Berne Three Step Test." Such a 'transformative use exception' would correspond very closely to the weak/no IP regime considered in the model presented here. Meanwhile in 2005 in the United States, the Supreme Court in *Merck KGaA v. Integra Life Sciences I, Ltd*[9] created a very broad research exemption in relation to pre-clinical R&D. Such a change again corresponds closely in the model to a move towards a weak/no IP regime in which a second-stage product would not infringe on a first-stage firm's patent.

2. A Basic Model of Two-Stage Cumulative Innovation

2.1. The Model.

We adopt a simple model of two stage innovation in which the second innovation builds upon the first in some manner – either as an application or as an extension of it. All agents are risk-neutral and act to maximize profits.

Innovations are described by their net value v (revenue minus costs). Because our interest lies in examining the trade-off between innovation at different stages we make no distinction between social and private value (i.e. there are no deadweight losses) and v may be taken to be both.

We assume the base (first) innovation takes two values: low (v_1^L) and high (v_1^H) with probability $p, (1-p)$ respectively. We assume that $v_1^L < 0$ so that without some additional source of revenue, for example from licensing (see below), the innovation will not be produced. High value innovations have positive stand-alone value, $v_1^H > 0$, and so do not require an outside source of revenue in order to be profitable.

Second-stage innovations also take two values: low (v_2^L) and high (v_2^H) with probability $q, (1-q)$ respectively and $v_2^H > v_2^L > 0$. While the value of a second-stage innovation is known to the innovator who produces it, the value is **not** known to the owner of the first-stage innovation which it builds upon (this could occur because of imperfect information regarding revenue, costs or both). Without loss of generality we shall assume that the number (or measure) of second-stage innovations per first-stage innovation is one (having N second-stage innovations per first-stage innovation would just require replacing v_2^H with Nv_2^H and v_2^L with Nv_2^L). We also assume that $v_1^L + v_2^L \geq 0$ – this ensures that whatever

[9]The full opinion is available at http://www.supremecourtus.gov/opinions/04slipopinion.html.

the value of q the overall value generated by a first-stage innovation is positive (the overall value is the stand-alone plus the value of dependent second-stage innovations).[10]

2.1.1. *Intellectual Property Rights and Licensing.* We wish to consider two regimes: one with (strong) intellectual property rights (IP) and one with weak, or no intellectual property rights (NIP). With intellectual property rights every second-stage innovator will require a license from the relevant first-stage innovator in order to market her product, while without intellectual property rights she may market freely without payment or licence.[11]

We assume that the direct returns to the first innovator (v_1) are unaffected by the intellectual property rights regime. This assumption is not as strong as it first appears since simple business stealing, in which the total combined rents of the two stages remain unchanged, could be incorporated into this model simply by increasing p, the proportion of first stage innovations that are low value.[12] Of course, if there is rent dissipation, due, say to further product market competition, this would not be the case and a richer model would be required. Given our need to keep the analysis tractable, and that the focus in this paper is on the division of rents between first and second-stage innovators, we do not take this approach, though we do return to the matter briefly in the conclusion.

Finally, we take the licence to define a lump-sum royalty payment r. This assumption is without loss of generality since, in this model, an innovation is entirely defined by its net value v and there are no other attributes available to use in designing a mechanism to discriminate between types of second-stage innovator.[13] The royalty is set ex-ante, that is prior to the second-stage innovator's decision to invest, and is in the form of a take-it-or-leave it offer by the first-stage innovator.

2.1.2. *Sequence of Actions.* The sequence of actions in the model is:

(1) Nature determines the value type of the first-stage innovator.

[10] Allowing values of v_1^L less than v_2^L does not alter the analysis in any significant way but brings extra complexity to the statement and proof of propositions.

[11] Given that we are dealing with cumulative innovation some readers might prefer the infringing (I) vs. non-infringing (NI) dichotomy with its implication of a distinction between 'horizontal' imitation and 'vertical' improvement of a product.

[12] The assumption would also be valid in the case where there is little substitution between the first and second-stage innovation. For example, where the first innovation is a tool used in developing the second-stage innovation.

[13] For example, there are no quantities on which to base a non-linear pricing scheme (fixed fee plus per unit fee royalty). For the same reason there is no opportunity to use type-contingent menus, or any other form of more complex licensing agreement, to increase total royalty income by discriminating between high and low value innovators.

Player	**Second-Stage Innovator**					
First Stage Innovator	**Value Type**		**Low (q)**		**High (1-q)**	
		Action	**NI**	**I**	**NI**	**I**
	Low (p)	**r**	v_1^L , 0	$v_1^L + r$, $v_2^L - r$	v_1^L , 0	$v_1^L + r$, $v_2^H - r$
	High (1-p)		v_1^H , 0	$v_1^H + r$, $v_2^L - r$	v_1^H , 0	$v_1^H + r$, $v_2^H - r$

Table 1. Action and Payoff Matrix Assuming First-Stage Innovator Invests. (I/NI = Invest/Do Not Invest, r = Royalty Rate)

(2) A first-stage innovator decide whether to invest. If the first-stage innovator does not invest the game ends and all payoffs are zero. Assuming the first-stage innovator invests the game continues.

(3) The first-stage innovator sets the royalty rate r (under the no/weak IP regime second-stage innovations do not infringe and so the de facto royalty rate is 0).

(4) Nature determines the value type of a second-stage innovator.

(5) Given this royalty rate second-stage firms decide whether to invest.

(6) Payoffs are realized.

The action/payoff matrix is summarized in Table 1.

2.2. **Solving the Model.** Define a constant, α, as follows:

$$\alpha \equiv \frac{v_2^H - v_2^L}{v_2^H}$$

Proposition 2.1. *With intellectual property rights, the game defined above has the following Subgame Perfect Nash equilibria. A second-stage innovator invests if and only if its realized value is greater than or equal to the royalty rate (i.e. net profits are non-negative). A first-stage innovator invests and sets a low royalty rate (RL), $r_L = v_2^L$ if the probability of a low value innovation (q) is greater than α and a high royalty rate (RH) $r_H = v_2^H$ if $q \leq \alpha$. When $q = \alpha$ the first-stage innovator may set any royalty of the form r_L with probability x and r_H with probability $1 - x$, $x \in [0, 1]$. Thus, there always exist a pure strategy equilibrium and except when $q = \alpha$, this equilibrium is unique.*

Proof. See appendix. □

Proposition 2.2. *Without intellectual property rights the game above has the following solution: both types of second-stage innovators invest but, of first-stage innovators, only those that have 'high-value' innovations invest (there are $1 - p$ of these type).*

	RL	**RH**
IP	$v_1 + v_2$	$v_1 + (1-q)(v_2^H)$
NIP	$(1-p)(v_1^H + v_2)$	$(1-p)(v_1^H + v_2)$
IP - NIP	$p(v_1^L + v_2^L) + p(v_2 - r_L) \geq 0$	$p(v_1^L + (1-q)v_2^H) - (1-p)qv_2^L$

Table 2. Welfare in the Basic Model

Proof. Trivial. □

2.3. **Welfare.** To determine welfare we need to know the 'trade-off' between first and second-stage innovations that occurs when revenue is allocated from one to the other by licensing. As stated above, without royalty income from second-stage innovations a proportion p of first-stage innovations are not produced with average (stand-alone) value v_1^L. The remaining innovations $(1-p)$ are produced irrespective of whether royalty revenue is received and have average value v_1^H.

Let us now consider social welfare in the four possible situations given by (IP, RL), (IP, RH), (NIP, RL), (NIP,RH) as well as the difference in welfare between an intellectual property regime and a no intellectual property regime (IP-NIP). Due to our earlier assumption welfare is determined by calculating total net value. Define for convenience $v_1 = pv_1^L + (1-p)v_1^H$, the average first-stage innovator value (if all innovate), and $v_2 = qv_2^L + (1-q)v_2^H$, the average second-stage innovator value (if all innovate). We summarize the welfare situation in Table 2.

2.4. **Policy Implications.**

Proposition 2.3. *When a low royalty will be set ($q \geq \alpha$) an IP regime is optimal.*

Proof. In the low royalty (RL) situation all second-stage innovations will be produced whether there is IP or not. In that case one wishes to maximize returns to the first innovator and patents do this by transferring rents via licensing. Formally in the low royalty case the welfare difference between patents and no patents (IP-NIP) is:

$$p(v_1^L + r_L) + p(v_2 - r_L)$$

Both of the terms in brackets are positive implying that the intellectual property regime delivers higher welfare than the no intellectual property (NIP) regime. □

The situation when the high royalty will be set is less clear. First, define β as the proportion of the royalty payment to a *low-value* first-stage innovator that would be 'used up' in paying their extra costs:

$$\beta \equiv \frac{-v_1^L}{(1-q)r_H}$$

Note that v_1^L is negative and must be less in absolute terms than the royalty received $(1-q)r_H$ as we are assuming that the royalty enables low value first-stage innovators to produce. Under this definition $\beta = 1$ corresponds to the case where all of the royalty paid to a low-cost first-stage innovator being used to pay their 'extra' costs while $\beta \approx 0$ means all of the royalty payment is being retained as extra profits (and welfare).

Proposition 2.4. *When a high royalty will be set ($q < \alpha$) an intellectual property regime will be preferable to a no intellectual property (NIP) regime if and only if (NB: in fact with equality one would be indifferent):*

$$p \geq \frac{qv_2^L + qv_2^L}{(1-\beta)(1-q)v_2^H} \tag{2.1}$$

$$= \frac{\textit{Licensing Failure Cost}}{\textit{Licensing Failure Cost} + \textit{Surplus From Extra 1st Stage}} \tag{2.2}$$

Proof. From Table 2 an IP regime yields higher welfare than an NIP regime if and only if:

$$p(v_1^L + (1-q)v_2^H) \geq (1-p)qv_2^L$$

Making p the subject of this inequality and using β we obtain the stated result. □

We represent the import of these propositions graphically in Figure 1, a diagram which shows optimal policy regions as a function of the exogenous probabilities of low value first-stage (p) and second-stage (q) innovations.

Remarks: in the high royalty case (RH) q is the proportion of second-stage innovations that do **not** occur **with** intellectual property rights (due to high royalties and the resulting licensing failure) while p is the proportion of first-stage innovations that do **not** occur **without** intellectual property rights. As first-stage innovations enable second-stage ones when we lose a first-stage innovation we lose all dependent second-stage ones as well. Due to this, when β is low for no intellectual property rights to be preferable q must be substantially higher than p. It is only then that the cost of intellectual property rights, in

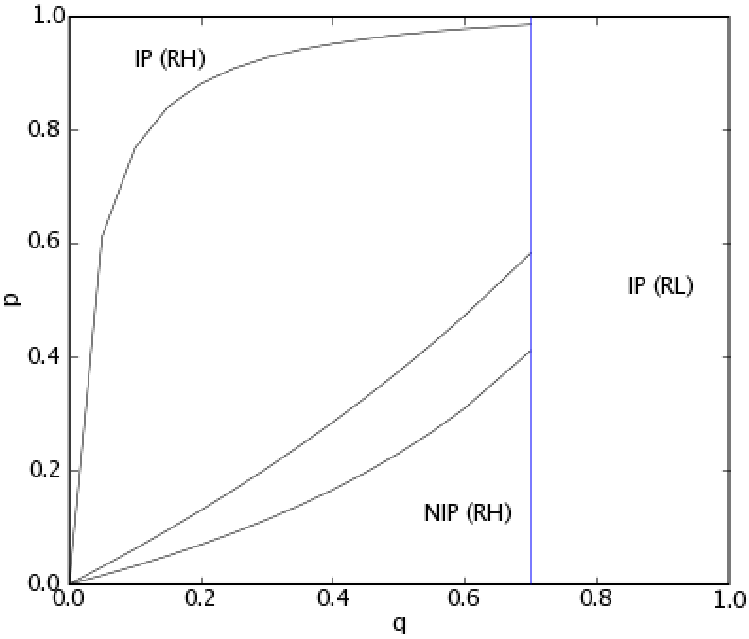

FIGURE 1. Optimal policy as a function of the probabilities of low value first-stage (p) and second-stage (q) innovations. α equals 0.7 so to the right of the line $q = 0.7$ a low royalty will be set (RL) and an IP regime is optimal. To the left of that line we have shown three different 'horizontal' lines which demarcate the boundary between IP being optimal (above the line) and no IP being optimal (below the line). The horizontal lines correspond (in ascending order) to β (the proportion of royalties used up by first stage-innovators) equal to 0, 0.5, and 0.99.

terms of lost second-stage innovations, will outweigh the gains in terms of more first-stage (and dependent second-stage) innovations.

As β increases the area in which no intellectual property rights are preferable will increase, with the line separating the two regions moving upwards. In the limit as β tends to 1 – which corresponds to all royalty income being used by a low value first-stage innovator to pay costs – the marginal p tends to 1, that is, it is optimal to have intellectual property rights only if all first-stage innovations are of a low value type.

3. A Model of Cumulative Innovation with Sampling

3.1. **The Model.** The 'sampling' model differs from the 'basic' model presented in the previous section only in the addition of a single extra period in which sampling by second-stage firms takes place prior to any royalty setting. Formally, we have the following modified sequence of actions (modifications are bolded for clarity):

(1) Nature determines the value type of the first-stage innovator.
(2) A first-stage innovator decide whether to invest. If the first-stage innovator does not invest the game ends and all payoffs are zero. Assuming the first-stage innovator invests the game continues.
(3) **Second-stage innovators chooses their level of sampling k.** (One could think of this, for example, as the number of first-stage products a second-stage firms chooses to investigate via purchase, observation etc).
    - Sampling has constant marginal cost τ.
    - Knowledge of the sampling level chosen by a second-stage firm. There are two possibilities regarding the knowledge of the sampling level available to first-stage innovators. In the first case the first-stage innovator does observe the sampling level. In the second case the first-stage innovator does not observe the sampling level. In what follows we focus on the case where the sampling level is unobserved as we feel this is more realistic though the results are unchanged (and simpler to derive) when it is observed.
(4) The first-stage innovator sets the royalty rate r (under the no/weak IP regime second-stage innovations do not infringe and so the de facto royalty rate is 0).
(5) Nature determines the value type of a second-stage innovator. As before there are two types of stage 2 firms, high and low value: v_2^H, v_2^L. However, here:
    - **The probability, q, that a second-stage firm is low value is a function of the sampling level: $q \equiv q(k)$.**
    - Properties of $q(k)$: $q' \leq 0$ (otherwise there is no benefit from sampling). There are diminishing returns to sampling: $q^{''} \geq 0$ and if no sampling takes place all firms are of low value type ($q(0) = 1$). The functional form $q(k)$ is assumed to be common knowledge.
(6) Given this royalty rate second-stage firms decide whether to invest.
(7) Payoffs are realized.

The new action/payoff matrix is shown in Table 3.

3.2. **Solving the Model.** Define, as in the basic model, a high royalty to be equal to the value of a high-value second-stage innovation: $r_H = v_2^H$, and a low royalty to be equal to the value of a low-value second-stage innovation: $r_L = v_2^L$.

Player			**Second-Stage Innovator**			
			Sample (k)			
First-Stage Innovator	**Value Type**		**Low (q(k))**		**High (1-q(k))**	
		Action	**NI**	**I**	**NI**	**I**
	Low (p)	**r**	v_1^L , $-k\tau$	$v_1^L + r$, $v_2^L - r - k\tau$	v_1^L , $-k\tau$	$v_1^L + r$, $v_2^H - r - k\tau$
	High (1-p)		v_1^H , $-k\tau$	$v_1^H + r$, $v_2^L - r - k\tau$	v_1^H , $-k\tau$	$v_1^H + r$, $v_2^H - r - k\tau$

TABLE 3. Action and Payoff Matrix Assuming First-Stage Innovator Invests (I/NI = Invest/Do Not Invest, r = Royalty Rate)

We begin with a set of preliminary propositions which detail the players best responses before moving on to characterise the equilibrium under both (strong) IP and weak/no IP (NIP).

Proposition 3.1 (Second-stage innovator's investment strategies)**.** *A second-stage innovator with value v_X facing a royalty of r will invest if and only if $v_X \geq r$.*

Proof. Just as in the original model second-stage innovator's move with full knowledge of all variables. In this case an innovator of type X invests if and only if net profits from investing, $v_X - r - k\tau$ are greater than $-k\tau$ the payoff from not investing (sampling costs are sunk). Hence the investment strategies are the same as in the basic model: a second-stage innovator invests if and only if $v_X \geq r$. □

Proposition 3.2 (First-stage Best-Response Royalty)**.** *Under the IP regime, a first-stage innovator, whose belief about the sampling level is given by the cdf $F(k)$ and where $\bar{q} = \mathbb{E}_F(q(k))$, will set a royalty of the form:*

$$r(k) = \begin{cases} r_L = v_2^L, & \bar{q} > \alpha \\ r_H = v_2^H, & \bar{q} < \alpha \\ \text{mixed strategy } (r_H, r_L) \text{ with prob } (x, 1-x), x \in [0,1], & \bar{q} = \alpha \end{cases}$$

where α is as in the basic model, that is the probability such that a first-stage firm is indifferent between setting a high and a low royalty rate:

$$\alpha \equiv \frac{v_2^H - v_2^L}{v_2^H}$$

Proof. See appendix. □

Remark 3.3 (Definition of k_α). If a first-stage innovator believes second-stage innovators all play the same pure strategy, k, then we can replace the conditions of the form $\bar{q} <, =, > \alpha$

with the condition that $k >, =, < k_\alpha$ (note the inversion of ordering), where the constant k_α, is the sampling level such that $q(k_\alpha) = \alpha$.

Proposition 3.4 (Second-Stage Sampling Level)**.** *Under an IP regime the second-stage innovators best response to a royalty of r, including 'composite' royalties of the form $r = xv_2^H + (1-x)v_2^L, x \in [0,1]$ (that is mixed royalty with r_H played with probability x), is as follows:*

$$k = \begin{cases} k_2, & r \leq r_L = v_2^L \\ k_r, & r_L < r < r_H \\ 0, & r \geq r_H = v_2^H \end{cases}$$

where k_r is defined implicitly by:[14]

$$q'(k_r) = \frac{-\tau}{v_2^H - r}$$

And k_2 is given as follows:[15]

$$k_2 = k_{r_L} = k_{v_2^L} \Rightarrow q'(k_2) = \frac{-\tau}{v_2^H - v_2^L}$$

Proof. See appendix. □

Theorem 3.5. *With intellectual property rights (IP) the perfect Bayesian equilibrium of the game defined above falls into one of two cases:*

(i) ***Low royalty case*** *($k_2 \leq k_\alpha$)*
   (1) *First-stage innovators: both high and low value types invest, believe that second-stage innovators sample at level k_2 and set a low royalty rate.*
   (2) *Second-stage innovators: sample at level k_2 and both high and low value types invest.*

(ii) ***Mixed royalty case*** *($k_2 > k_\alpha$)*
   (1) *First-stage innovators: both high and low value types invest, believe that second-stage innovators sample at level k_α and set a mixed royalty rate consisting of a high royalty (r_H) with probability x_α and a low royalty (r_L) with*

[14] If $q'(0) > -\infty$ then for values of r sufficiently close to $r_H = v_2^H$ this equation will have no solution. In such cases define $k_r = 0$.

[15] We use the subscript 2 because this is the level of sampling undertaken when all second-stage innovators (both high and low types) always.

probability $(1 - x_\alpha)$ *where:*[16]

$$x_\alpha = 1 - \frac{\tau}{-q'(k_\alpha)(v_2^H - v_2^L)}$$

(2) *Second-stage innovators: sample at level* k_α *and invest if and only if the realized value of their innovation is greater than the royalty rate (though the first-stage innovator is playing a mixed strategy the second-stage innovator knows the royalty rate with certainty at the point of investment).*

Proof. See appendix. □

Proposition 3.6 (Equilibrium under weak/no IP)**.** *Under weak/no IP the 'sampling' model has the following solution: second-stage innovators sample at level* k_2 *and both types of second-stage innovators invest. Of first-stage innovators, those that have 'high-value' innovations invest (there are* $1 - p$ *of these type) and those with 'low-value' innovations do not.*

Proof. Trivial. (Second-stage sampling best-response correspondences have already been derived in Proposition 3.4). □

Remark 3.7. Recall that k_2 is the sampling level undertaken by a second-stage firm in the case when both high and low value second-stage innovators invest (so it occurs either in the case where there is no IP or when the royalty is sufficiently low). It is also, therefore, the sampling level which maximizes expected second-stage innovation value and, for that reason, the socially optimal sampling level.

3.3. **Welfare.** For the welfare calculations we proceed as in the original model. A proportion p of first-stage innovations are low value ($v_1^L < 0$) and only occur when there is royalty income. Analogously to the basic model define $v_1 = pv_1^L + (1-p)v_1^H$ and $v_2(k) = -k\tau + (1 - q(k))v_2^H + q(k)v_2^L$ (the expected value generated by a second-stage innovator sampling at level k).

Proposition 3.8. *[The Optimal Regime in the Low Royalty Case] In the low royalty case* ($k_2 < k_\alpha$) *it is optimal to have an IP regime (compared to weak/no IP one). Specifically if*

[16]Note examining the definition of k_2 shows that $k_\alpha < k_2$ guarantees that x_α is non-negative.

the proportion (p) of first-stage innovation that is lost without IP is positive then welfare is higher with IP (otherwise $p = 0$ and both regimes generate the same level of welfare).

Proof. See appendix. □

This result has a simple intuition behind it. The low royalty case encompasses the situation where the sampling level is fairly low even when the royalty rate faced by second-stage firms is small ($k_2 \leq k_\alpha$) – this may occur because sampling is costly (τ is high) or generates little benefit ($v_2^H and v_2^L$ are close). As a result most second-stage innovations are low value and so a first-stage innovator sets a low royalty rate (r_L). Hence (a) there is no 'licensing failure' and (b) all second-stage firms sample at the optimum rate (k_2). Taken together these mean that, just as with the low royalty case of the simpler model, there are no costs to having strong IP. Since, thanks to the licensing income, there is more (by an amount p) first stage innovation under strong IP than under weak/no IP the strong IP regime is clearly better.

Proposition 3.9. *[The Optimal Regime in the Mixed Royalty Case] In the mixed royalty case ($k_2 \geq k_\alpha$) it is optimal to have an IP regime rather than a weak/no (NIP) regime if the proportion (p) of first-stage innovation that does* not *occur under no/weak IP is sufficiently high, specifically:*

$$p \geq p^m \equiv \frac{(v_2(k_2) - v_2(k_\alpha)) + x_\alpha q(k_\alpha) v_2^L}{(v_2(k_2) - v_2(k_\alpha)) + x_\alpha q(k_\alpha) v_2^L + ((v_2(k_\alpha) - x_\alpha q(k_\alpha) v_2^L - (-v_1^L))} \tag{3.1}$$

$$= \frac{\textit{Reduced Sampling Cost} + \textit{Licensing Failure Cost}}{\textit{Reduced Sampling Cost} + \textit{Licensing Failure Cost} + \textit{Surplus from Extra 1st Stage}} \tag{3.2}$$

Proof. See appendix. □

Remark 3.10. Reduced Sampling Cost: $v_2(k_2)$ is the average value of second-stage innovations when second-stage firms sample at the unrestricted (and optimal) level k_2. Under the IP regime second-stage firms only sample at level k_α because of the higher (average) royalty. Thus, the average value of a second-stage innovation is less under the IP regime compared to the weak/no IP regime due to this reduced sampling precisely by the amount:

$v_2(k_2) - v_2(k_\alpha)$ (NB: obviously this only applies to those second-stage innovations associated with the $(1-p)$ first-stage innovations which are produced under both the IP and the weak/no IP regime.)

Licensing Failure Cost: licensing failure occurs when a second-stage firm with a low-value innovation is faced with a high royalty rate. Under the IP regime x_α is the probability that a high royalty is set by a first-stage innovator $q(k_\alpha$ is the probability a second-stage firm has a low-value innovation. Thus $x_\alpha q(k_\alpha)$ is the probability that licensing failure occurs and when it does the loss equals the potential value of the second-stage innovation: v_2^L.

Surplus from Extra First-Stage Innovation: the plus side of the IP regime is the extra first (and dependent) second-stage innovation that happens because first-stage innovators receive higher incomes. There are a proportion p of low (standalone) value first-stage innovators, who will only invest under the (strong) IP regime. For each such innovation the net surplus generated equals the surplus generated by the second-stage firms plus the net (stand-alone) surplus of a first-stage firm. The expected second-stage surplus equals the average value if all second-stage firms produced (when sampling at k_α: $v_2(k_\alpha)$, minus the surplus of those second-stage firms who are held-up: $x_\alpha q(k_\alpha) v_2^L$. Finally the net standalone surplus of a first-stage firm is $v_1^L < 0$.

Finally, compare equation (3.1) with equation (2.1) from the basic model. The main, and most obvious, difference is that, as well as the standard 'licensing failure cost' of (strong) IP, there is another, additional, cost in the form 'reduced sampling' (and reduced average value of second-stage innovations).

Corollary 3.11. *Extending $p^m = 0$ to the low royalty case ($k_2 \leq k_\alpha$) by defining $p^m = 0$ if $k_2 \leq k_\alpha$, we have that an IP regime is optimal if $p > p^m$ and a weak/no IP is optimal if $p < p^m$.*

3.4. **Policy Implications.** Since we do not have any precise estimates for the exogenous parameters such as the sampling cost (τ) or the values of second-stage innovations (v_2^H etc) we cannot make direct statements about which regime would yield higher welfare for a given industry. Instead our approach has been to to pick a 'dependent' variable to focus on (in our case p, the proportion of first-stage innovation 'lost' under weak/no IP)

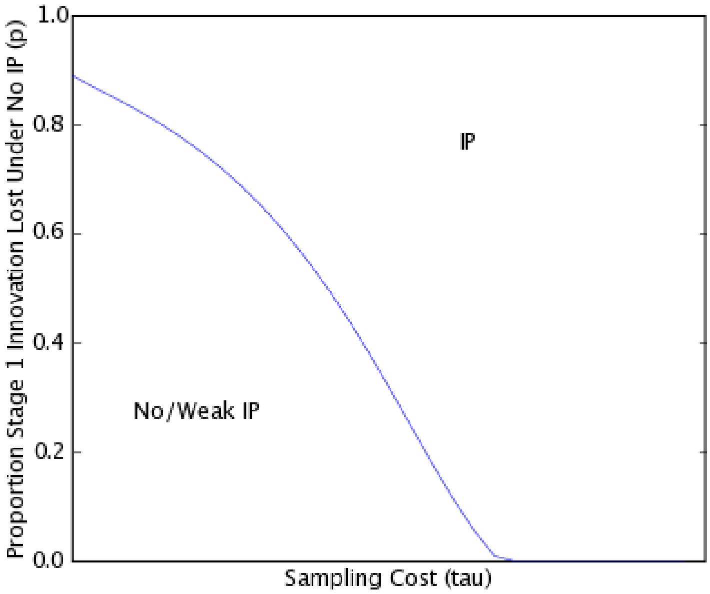

FIGURE 2. Marginal p^m as a function of sampling cost. Ticks on τ axis have been specifically omitted as they would be misleading – any particular value will depend on parametrization, functional form for q etc. However in this specific case we note that $v_2^H = 10, v_2^L = 1.0, q(k) = e^{-k}$ and $p^m = 0$ above 8.0.

and then derive the 'break-even' or marginal p^m such that if $p = p^m$ we are indifferent in welfare terms between the two regimes.

Our next step is to investigate the comparative statics of the marginal p (p^m) with respect to exogenous variables, in particular the cost of sampling (τ) and the relative value of high (v_2^H) and low type (v_2^L) second-stage innovations.

Our general results are summarized in Figure 2 and Figure 3. As we note in the captions one can only indicate the general form as any specific form for p^m will depend on the functional form for q and of course the values of the other exogenous parameters.

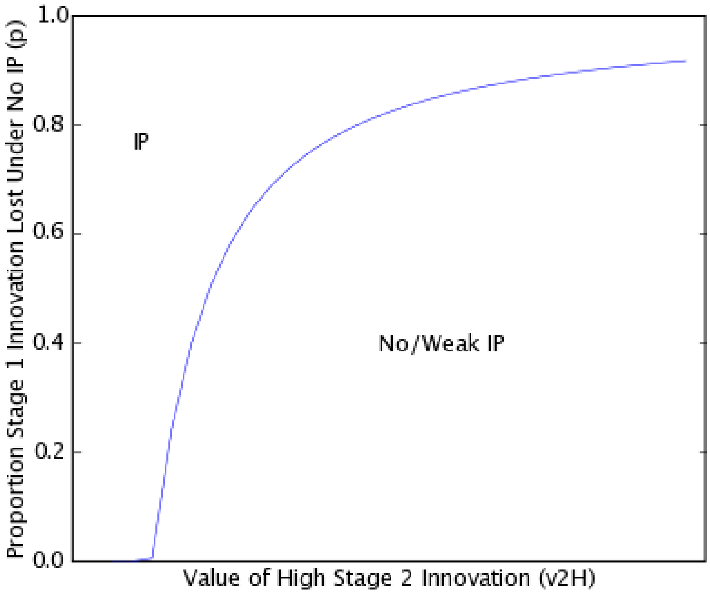

FIGURE 3. Marginal p^m as a function of v_2^H (or equivalently, for fixed v_2^L: $v_2^L - v_2^L$). For the same reasons given in relation to Figure 2 ticks on the v_2^H axis have been omitted. However to give the reader some sense of proportion we note that $\tau = 0.5, v_2^L = 1.0, q(k) = e^{-k}$ and $p^m = 0$ below approximately 3.0.

Proposition 3.12. *The sampling levels k_α and k_2 have the following comparative statics:*

$$\frac{dk_\alpha}{d\tau} = 0 \tag{3.3}$$

$$\frac{dk_\alpha}{dv_2^H} < 0 \tag{3.4}$$

$$\frac{dk_2}{d\tau} < 0 \tag{3.5}$$

$$\frac{dk_2}{dv_2^H} > 0 \tag{3.6}$$

And taking limits:

$$\lim_{\tau \to \infty} k_2 = 0, \lim_{v_2^H \to v_2^{L+}} k_2 = 0, \lim_{\tau \to 0} k_2 = \infty, \lim_{v_2^H \to \infty} k_2 = \infty \tag{3.7}$$

$$\lim_{v_2^H \to \infty} k_\alpha = 0 \tag{3.8}$$

Proof. Recall that we have:

$$q(k_\alpha) = \frac{v_2^H - v_2^L}{v_2^H}$$
$$q'(k_2) = \frac{-\tau}{v_2^H - v_2^L}$$

Given that $q' < 0$ and $q'' > 0$ the results following trivially by simple differentiation. □

Remark 3.13. The intuition behind these results is straightforward. k_α is the level of sampling that leaves a first-stage innovator indifferent between charging a high and a low royalty rate. As such it is a function only of the relative values of the two types of innovation (and of q) and does not depend on the cost of sampling at all.

The intuition in the second case is a little more complicated. If we increase v_2^H keeping v_2^L constant we increase the differential between high and low value second-stage innovations. Then the net change in revenue for a first-stage innovator's from switching to a high royalty rate must increase (loss of royalty revenue from low-value second-stage innovations is lower relative to royalty from high-value second-stage innovations). Hence, the proportion of high value second-stage innovations $(1 - q(k))$ at which the switch to a high royalty rate is made is smaller and the corresponding level of sampling (k_α) is smaller.

Coming to k_2, which is the optimal level of sampling (and that performed under a low or zero royalty), we have unsurprisingly that as the cost of sampling goes down the amount of sampling goes up. Similarly, an increase in the relative size of a high value innovation compared to a low value one, increases the benefit of sampling and therefore increases the amount of sampling done.

Combining the differentials with the limits we have that (a) keeping other variables fixed there exists a unique finite τ^* such that for $\tau < \tau^*, k_2 > k_\alpha$ and a mixed royalty is set (conversely for $\tau > \tau^*$ a low royalty is set and $p^m = 0$); (b) similarly there exists a unique v^* such that for $v_2^H > v^*, k_2 > k_\alpha$ and a mixed royalty is set (conversely for $v_2^H < v_2^{H*}$ a low royalty is set and $p^m = 0$). This then demonstrates the validity of the right-hand part of Figure 2 and the left-hand part of Figure 3 where we have $p^m = 0$.

What occurs then if $k_2 > k_\alpha$ and we are in the mixed royalty case?

Proposition 3.14. *Assuming $k_2 > k_\alpha$ (i.e. τ sufficiently small or v_2^H sufficiently large) then:*

$$p \geq p^m \equiv \frac{(v_2(k_2) - v_2(k_\alpha)) + x_\alpha q(k_\alpha) v_2^L}{(v_2(k_2) - v_2(k_\alpha)) + x_\alpha q(k_\alpha) v_2^L + ((v_2(k_\alpha) - x_\alpha q(k_\alpha) v_2^L) - (-v_1^L))}$$

And we have that:

$$\frac{dp^m}{d\tau} < 0$$
$$\frac{dp^m}{dv_2^H} > 0$$

That is, the marginal level of first-stage innovation lost under weak/no IP (that is the level such that above this an IP regime is optimal) is (a) decreasing in sampling costs (b) increasing in the relative size of high value to low value second-stage innovations.

Proof. See appendix. □

Informally this result can be explained as follows. Reductions in sampling costs will increase the ‘optimal’ level of sampling (k_2) relative to the restricted level of sampling (k_α). This in turn increases the cost of intellectual property rights arising from (a) loss of second-stage innovations due to licensing failure ($x_\alpha \cdot q(k_\alpha)$); (b) lower average value of second-stage innovations ($v_2(k_2) - v_2(k_\alpha)$); while having no effect on the surplus from extra first-stage innovations under IP. As a result the welfare under weak/no IP rises relative to the welfare under IP and the marginal p must rise.

Similarly if the relative size of high value second-stage innovation compared to a low value one rises this (a) increases the ‘optimal’ level of sampling (k_2) relative to the restricted level of sampling (k_α) (b) directly increases the benefit of sampling. This again increases the sampling cost and the licensing failure cost but reduces the surplus from second-stage innovations under IP. As a result welfare under weak/no IP rises relative to that under IP and the marginal p must rise.

This result then establishes the validity of the rest of Figures 2 and 3 and implies the following corollaries regarding how the optimal policy regime in relation to intellectual property rights varies in response to changes in the exogenous environment:

Corollary 3.15. *Reducing sampling costs make it more likely that a freer (weak/no intellectual property rights) regime will be optimal.*

Proof. Follows from previous propositions as summarised in Figure 2. □

Corollary 3.16. *Increasing the differential between high and low value second-stage innovations (which could be interpreted as sampling becoming more important for product quality) makes it more likely that a freer (no intellectual property rights) regime will be optimal.*

Proof. Follows from previous propositions as summarised in Figure 3. □

Remark 3.17. Most studies of the value of intellectual property rights (copyrights or patents) indicate that their distribution is highly skewed with a few very high value works and many low value works. This suggests that $v_2^H >> v_2^L$.

4. Conclusion

In this paper we have shown how asymmetric information about the value of follow-on innovations, combined with intellectual property rights such as patents, can result in licensing failure and hold-up. Presenting the policy decision as a choice between having or not having intellectual property rights, we have shown that, in contrast to parts of the previous literature, in some circumstances it may be optimal not to have intellectual property rights. For whilst intellectual property rights help transfer income from second-stage to first-stage innovators they can also lead to licensing failure and hold-up with a resulting reduction in second-stage innovation.

In the first, and simpler, model presented, the basic results were summarized in Figure 1, which plotted optimal policy as a function of the exogenous variables (the probabilities of high or low value innovations occurring at the two different innovation stages). Intellectual property rights in this model had two contrasting effects. On the one hand, there are the benefits of increased first-stage innovation as revenue is transferred to first-stage innovators from second-stage ones. On the other hand, there are costs in terms of fewer second-stage innovations due to licensing failure. In some circumstances the benefits will exceed the costs and we should have (stronger) intellectual property rights. In other cases, they will not and we should have weaker (or no) intellectual property rights. In particular, we showed that, if the probability of a low value second-stage innovation was high enough (but not too high), compared to the probability of a low value first-stage innovation, then a regime without intellectual property rights would be preferable.

Next, we extended this basic model by introducing 'sampling'. We demonstrated the existence of a perfect Bayesian equilibrium and showed that (strong) IP may restrict the level of sampling below what would be socially optimal. Therefore, in addition to the basic trade-off mentioned above between more first-stage innovations and fewer second-stage ones, there is the additional factor: those second-stage innovations which occur have lower average value due to a lower level of sampling. Examining this trade-off, we find that the lower the cost of sampling and the greater the differential between the low and high values of second-stage innovations, the more likely it is that a regime *without* intellectual property rights will be preferable.

Thus, technological change which reduces the cost of encountering and trialling new 'ideas' should imply a reduction in the socially optimal level of intellectual property rights such as patents and copyright. A perfect case of such technological change in recent years can be found in the rapid advances in computers and communications. These advances have, for example, dramatically reduced the cost of accessing and re-using cultural material, such as music and film, as well as greatly increasing the number of 'ideas' that a software developer can encounter and trial. Concrete policy actions that could be taken in line with these conclusions include extending 'fair-use' (fair-dealing) provisions in copyright law to increase the degree of reuse that would be permitted without the need to seek permission and excluding software and business methods from patentability.

Finally, we should emphasize that there remains plentiful scope to improve and extend the present paper. For instance, it was assumed that the non-royalty income for the first-stage and second-stage innovator was unaffected by the intellectual property rights regime.[17] However this is unlikely to be the case and the model could be improved by the inclusion of the direct effect of no (or weaker) intellectual property rights on the revenue of the first-stage (and second-stage) innovator.

It would also be useful to extend the analysis to the case of a continuous distribution of innovation values, as well as to investigate the consequences of making sampling costs

[17] As discussed in detail above, while we do allow for business stealing between the first and second-stage innovators we do not allow for general rent dissipation from wider product market competition.

a function of the intellectual property rights regime. It would also be valuable to examine what occurs when the structure of innovation is more complex, for example by having second-stage inventions incorporate many first-stage innovations (a componentized model) or having heterogeneity across innovations with some developments used more than others. Finally, one of the most important extensions would be to properly integrate transaction costs into the analysis. Transaction costs relating to both the acquisition of information and the execution of contracts are significant and without them we lack a key element for the furtherance of our understanding of the process of innovation both in this model and in general.

A. Proofs

A.1. **Proof of Proposition 2.1.**

Proof. We are considering only subgame perfect nash equilibria so we may begin at the final stage of the game and work backwards. Given a royalty level of r, at the final stage, a second-stage innovator of type X faces a payoff of $v_2^X - r$ if she invests and 0 if she does not. Thus, a second-stage innovator, seeking to maximize profits will invest if and only if $v_2^X \geq r$ (formally, they are indifferent if $r = v_2^X$. However if they do not invest when $v_2^X = 0$ there will be no equilibrium of the overall game).

Given this, by simple dominance and focusing on pure strategies, a first-stage innovator must EITHER (a) set a low royalty rate $r_L = v_2^L$ which will lead to investment by all second-stage innovations; OR (b) set a high royalty rate $r_H = v_2^H$ which will result in investment only by high value second-stage innovations. In the first case the payoff is r_L while in the second it is $(1-q)r_H$. Thus, a low royalty rate should be chosen if and only if (assuming that if payoffs are equal a low royalty is chosen):

$$r_L \geq (1-q)r_H \iff q \geq \frac{r_H - r_L}{r_H} = \alpha$$

Since any mixed royalty strategy must consist of some combination of r_L and r_H we have immediately that a proper mixed strategy is only possible when $r_L = (1-q)r_H$, that is if $q = \alpha$.

Finally, total royalty income to a first-stage innovator is at least $r_L = v_2^L$. Thus, total net income for a low-value first-stage innovator is at least $v_1^L + r_L = v_1^L + v_2^L > 0$ (by assumption) – and net income for a high-value first-stage innovator is obviously greater. Hence both types of first-stage innovator will invest. □

A.2. **Proof of Proposition 3.2.**

Proof. Given a first-stage innovator believes $F(k)$, the expected probability that a second-stage firm is low value is $\mathbb{E}_F(q(k)) = \bar{q}$. By subgame perfection a first-stage innovator knows that, once a seond-stage firm discovers its type, its best response to a given royalty will be as stated in Proposition 3.1. In particular, if the royalty rate is set to be less than or equal to the second-stage low value (v_2^L) all second-stage innovators will license, if a royalty is above this but less than or equal to the second-stage high value (v_2^H) then only high value firms will license ($1 - \bar{q}$ of them) and if the royalty is higher than this no second-stage firms will license. Then, letting $G(r)$ be the cumulative distribution function over royalties representing the first-stage innovator's mixed strategy, the expected payoff to a first-stage innovator is:

$$\Pi_1(G(r)) = \int_0^{v_2^L} r \cdot dG(r) + (1 - \bar{q}) \int_{v_2^L}^{v_2^H} r \cdot dG(r) + 0 \cdot \int_{v_2^H}^{\infty} r \cdot dG(r))$$

Maximizing with respect to $G(r)$ immediately gives that, just as for the basic model, an optimal mixed strategy can only consist of some combination of the pure strategy $r_L = v_2^L$ and the pure strategy $r_H = v_2^H$. Let us suppose that these two pure strategies, r_H, r_L, are played with probability $x, 1 - x$ respectively. Revenue from royalties is then:

$$r_L(1 - x) + (1 - \bar{q})r_H x = r_L + x \cdot ((1 - \bar{q})r_H - r_L)$$

Maximizing revenue requires $x = 0$ if the term in brackets is less than zero, $x = 1$ if the term in brackets is greater than 0, and allows any value of x if the term in brackets is zero. By the definition of α (see above) these conditions correspond precisely to $\bar{q}$ (the expected probability of a low value innovation) being less than, greater than or equal to α. Hence, the first-stage innovator's royalty response as a function of their belief about the level of sampling is of the form stated. □

A.3. **Proof of Proposition 3.4.**

Proof. Using the optimal investment stage determined in Proposition 3.1, for a given sampling level k, payoffs as a function of the royalty levels are as in Table 4.

	$r \leq r_L$	$r_L < r < r_H$	$r \geq r_H$
$\Pi(k)$	$-k\tau - r + q(k)v_2^L + (1-q(k))v_2^H$	$-k\tau + (1-q(k))(v_2^H - r)$	$-k\tau$

Table 4. Payoff for Second Stage Innovator

Suppose second-stage innovator plays a strategy given by the cdf $F(k)$ and a first-stage innovator sets a royalty defined by a cumulative distribution function $G(r)$. Then the payoff to a second-stage innovator is as follows (where expectations are taken with respect to F and q is short for $q(k)$):

$$\Pi_2(F(k)) = \mathbb{E}\left(-\tau k + \int_0^{r_L} qv_2^L + (1-q)v_2^H - rdG(r) + \int_{r_L}^{r_H} (1-q)(v_2^H - r)dG(r) + \int_{r_H}^{\infty} 0dG(r)\right)$$
$$= \mathbb{E}\left(-\tau k - q\{G(r_H)v_2^H - G(r_L)v_2^L - \int_{r_L}^{r_H} rdG(r)\} + G(r_H)v_2^H - \int_0^{r_H} rdG(r)\right)$$

Claim: Second-stage innovators play pure strategies.

Proof: q is convex so $-q$ is concave. Suppose we have a mixed strategy $F(k)$ with $\mathbb{E}(k) = \bar{k}$ then $-\bar{q} = \mathbb{E}(-q(k)) \leq -q(\bar{k})$ with equality if and only if $F(k)$ is a point distribution (i.e. corresponds to a pure strategy). Substituting:

$$\Pi_2(F(k)) = \mathbb{E}_F(-\tau k + G(r_H)v_2^H - q(G(r_H)v_2^H - G(r_L)v_2^L) - \int_0^{r_H} rdG(r)$$
$$= -\tau\bar{k} - \bar{q} \cdot (+\text{ve}) + \text{const}$$
$$\leq -\tau\bar{k} - q(\bar{k}) \cdot (+\text{ve}) + \text{const}$$

(With equality iff and only if $F(k)$ is a point distribution with $k = \bar{k}$ with probability 1). Thus for any properly mixed strategy $F(k)$ we can always achieve a higher payoff by playing the pure strategy $\bar{k} = \mathbb{E}(k)$. □

Thus, in what follows we may confine our attention to pure strategies k. Returning to the payoff function we first note that if royalty (or royalties in a mixed strategy) are all greater than r_H (formally the support of $G(r)$ lies entirely above r_H) then the optimal sampling level is zero ($\Pi_2(k) = -k\tau$).

When this is not the case we have the first order condition is:[18]

[18]The second order condition, $\Pi^{''} \leq 0$, is easily checked: $\Pi^{''} = -q^{''}(k) \cdot (+\text{ve}) < 0$ since, by assumption, $q^{''}(k) > 0$.

$$q'(k) = \frac{-\tau}{G(r_H)v_2^H - G(r_L)v_2^L - \int_{r_L}^{r_H} r dG(r)}$$

For ease of reference define S as the denominator in the previous equation. We shall look at several special cases as follows:

(i) $r \leq r_L$. Then $G(r_H) = G(r_L) = 1$ and we have $S = v_2^H - v_2^L$. The profit-maximizing k therefore equals k_2 where (as defined above):

$$q'(k_2) = \frac{-\tau}{v_2^H - v_2^L}$$

The intuition here is simple: both firms always invest and pay the royalty. Thus, in terms of the payoff sampling will only affect the value type and the sampling level will be chosen so that the marginal gain in terms of lower costs, $q'(k)(v_2^H - v_2^L)$, equals the marginal sampling costs, τ.

(ii) $r_L < r < r_H$. Here $G(r_L) = 0, G(r_H) = 1$ and we have $S = v_2^H - r$ and the optimal $k \equiv k_r$ solves:

$$q'(k_r) = \frac{-\tau}{v_2^H - r}$$

(iii) r_H played with probability x and r_L with probability $(1-x)$. Then $G(r_L) = (1-x), G(r_H) = 1$. Define the 'composite' royalty $r = xr_H + (1-x)r_L = xv_2^H + (1-x)v_2^L$ then we have $S = v_2^H - (1-x)v_2^L - xr_H = (1-x)(v_2^H - v_2^L) = v_2^H - r$. So the optimal sampling level is $k \equiv k_r$ where r is the composite royalty.

□

A.4. **Proof of Theorem 3.5.**

Proof. We will solve for a subgame perfect Bayesian nash equilibrium by recursing backwards through the game.

In previous propositions we have already derived the best-response correspondences (where the royalty best-response is defined in terms of *beliefs* about sampling rather than the actual sampling level). We have also shown second-stage firms will always play a pure strategy (i.e. choose a single sampling level). Furthermore, at the sampling stage all second-stage firms are the same, hence all second-stage firms will choose the same pure sampling strategy. Thus, a first-stage innovator's beliefs (to be consistent) must be single-valued and we may rewrite the royalty best-response correspondence in terms of their belief as to the sampling level (k):[19]

$$r(k) = \begin{cases} r_L = v_2^L, & k < k_\alpha \\ r_H = v_2^H, & k > k_\alpha \\ \text{mixed strategy } (r_H, r_L) \text{ with prob } (x, 1-x), x \in [0,1], & k = k_\alpha \end{cases}$$

Case 1: $k_2 \leq k_\alpha$. There are three possibilities for the beliefs of a first stage innovator regarding the sampling level of second-stage firms:

(i) $k > k_\alpha$. Hence the first-stage innovator would set a high royalty rate. Then second-stage innovator's best response is $k = 0$ and beliefs will be inconsistent. Thus, there cannot be an equilibrium with such beliefs.

(ii) $k < k_\alpha$. In this case the best response of a first-stage innovator is to set a low royalty (r_L) in which case second-stage firm must choose a sampling level $k = k_2$. Thus, for beliefs to be consistent, a first-stage innovator must believe $k = k_2$ and the equilibrium is as claimed.

(iii) $k = k_\alpha$. In this case a first-stage innovator's best response correspondence consists of all mixed strategies: r_H with probability x, r_L with probability $1-x$ for $x \in [0,1]$. Now a second-stage innovator (if behaving optimally) never samples above the level k_2 and will sample strictly below k_2 if the first-stage innovator plays any strategy in which r_H is played with positive probability. Hence if beliefs are to be consistent we must have (a) $k_2 = k_\alpha$ and (b) $x = 0$ (i.e. a low royalty is always set). In such a case the equilibrium is again as claimed.

Case 2: $k_2 > k_\alpha$. There are three possibilities for the beliefs of a first stage innovator regarding the sampling level of second-stage firms:

(i) $k > k_\alpha$. Just as in the first case this leads to inconsistent beliefs and so cannot be an equilibrium.

(ii) $k < k_\alpha$. In this case the best response of a first-stage innovator is to set a low royalty (r_L) in which case second-stage firm must choose a sampling level $k = k_2$. But $k_2 > k_\alpha$. Thus, beliefs will be inconsistent and this cannot be an equilibrium.

[19] At the sampling stage all second-stage firms are the same and their best-response correspondence is single-valued. Hence all second-stage firms must have the same sampling strategy and a first-stage innovator's belief

(iii) $k = k_\alpha$. In this case a first-stage innovator best response correspondence consists of all mixed strategies: r_H with probability x, r_L with probability $1-x$ for $x \in [0,1]$. Denote the corresponding composite royalty by $r(x) = xr_H + (1-x)r_L$. Then for an equilibrium (with consistent beliefs) we must find an x such that the best-response sampling level equals k_α. Formally, using the notation of Proposition 3.4 we must find an x such $k_{r(x)} = k_\alpha$. The best response sampling level is defined implicitly by:

$$q'(k) = \frac{-\tau}{(1-x)(v_2^H - v_2^L)}$$

Since $q' < 0$ we have, denoting $k(x)$ as the implicit solution as a function of x, that $k'(x) < 0$ (intuitively a higher average royalty lowers sampling). Since $k(0) = k_2 > k_\alpha$ and that $k(1) = 0$ (as $x \to 1$ the RHS of the above takes arbitrarily large negative values), by the intermediate value theorem and the monotonicity of $k(x)$, there must exist a unique $x_\alpha \in (0,1)$ such that $k(x_\alpha) = k_\alpha$. Replacing $q'(k)$ by $q'(k_\alpha)$ and rearranging we have as claimed that:

$$x_\alpha = 1 - \frac{\tau}{-q'(k_\alpha)(v_2^H - v_2^L)}$$

First-stage innovators investment strategy: finally as with our basic model first-stage innovators of both types invest because with royalty income net profits will be non-negative. □

A.5. **Proof of Proposition 3.8.**

Proof. Analogously to the low royalty case in the basic model, in this situation all second-stage innovators invest so (a) there is no licensing failure (b) second-stage firms sample at the optimal level (k_2). At the same time, intellectual property allows some first-stage innovators to engage in production who wouldn't be able to do so otherwise. Hence an IP regime will deliver higher welfare.

Formally, the welfare difference between the IP and NIP regime is net surplus associated with the p extra first-stage innovations that occur under IP:

$$p((v_1^L + r_L) + (v_2(k_2) - r_L))$$

Both the first term (by the assumption that the royalty is sufficient to allow production) and the second (since second-stage innovators are making non-negative profits) are positive. Hence, if $p > 0$ the sum is positive and welfare is higher with intellectual property. □

A.6. **Proof of Proposition 3.9.**

Proof. In this case comparing the IP to the no/weak IP regime we have the following differences:

(+) Under IP there are (p) extra first-stage (and dependent second-stage) innovation because the royalty income allows some first-stage innovators to produce who would not otherwise:

$$p \underbrace{(v_1^L + v_2(k_\alpha) - x_\alpha q(k_\alpha) v_2^L)}_{\text{surplus per extra first stage innovation}}$$

(-) For the $(1-p)$ first-stage innovations that occur under both IP and no/weak IP there are fewer associated second-stage innovations due to licensing failure (licensing failure cost) and the innovations are of lower average value due to reduced sampling (reduced sampling cost):

$$-(1-p)(\underbrace{(v_2(k_2) - (v_2(k_\alpha))}_{\text{Reduced Sampling Cost}} + \underbrace{x_\alpha q(k_\alpha) v_2^L}_{\text{Licensing Failure Cost}})$$

An IP regime is optimal compared to a weak/no IP (NIP) if the first effect is larger than the second (and vice versa):

$$p(v_1^L + v_2(k_\alpha) - x_\alpha q(k_\alpha) v_2^L) - (1-p)(v_2(k_2) - v_2(k_\alpha) + x_\alpha q(k_\alpha) v_2^L) \geq 0$$

$$\Leftrightarrow p \geq p^m \equiv \frac{(v_2(k_2) - v_2(k_\alpha)) + x_\alpha q(k_\alpha) v_2^L}{(v_2(k_2) - v_2(k_\alpha)) + x_\alpha q(k_\alpha) v_2^L + ((v_2(k_\alpha) - x_\alpha q(k_\alpha) v_2^L) - (-v_1^L))}$$

Where p^m has been defined as the probability of a low value first-stage innovation which leaves one indifferent between having and not having intellectual property rights. □

A.7. **Proof of Proposition 3.14.**

Proof. Define:

$$S = \text{Higher Sampling Cost} = (v_2(k_2) - v_2(k_\alpha))$$

$$H = \text{Licensing Failure Cost} = x_\alpha q(k_\alpha) v_2^L$$

$$E = \text{Surplus per Extra Stage 1} = v_2(k_\alpha) - x_\alpha q(k_\alpha) v_2^L - (-v_1^L)$$

Then,

$$p^m = \frac{S+H}{S+H+E}$$

Examining the differentials of S, H, E we have:

$$\begin{aligned}
\frac{dS}{d\tau} &= \frac{\partial}{\partial \tau}(v_2(k_2) - v_2(k_\alpha)) + \frac{dv_2(k_2)}{dk_2}\frac{dk_2}{d\tau} - \frac{dv_2(k_\alpha))}{dk_\alpha}\frac{dk_\alpha}{d\tau} \\
&= (-) + (+ \cdot -) + (+ \cdot 0) = - \\
\frac{dH}{d\tau} &= \frac{dx_\alpha}{d\tau}(\cdots) + (\cdots)\frac{dk_\alpha}{d\tau} = (- \cdot +) + 0 = - \\
\frac{dE}{d\tau} &= \frac{dE}{dk_\alpha}\frac{dk_\alpha}{d\tau} = (\cdots) \cdot 0 = 0
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{dS}{dv_2^H} &= + \\
\frac{dH}{dv_2^H} &= + \\
\frac{dE}{dv_2^H} &= -
\end{aligned}$$

For the last equation note, that by definition of k_α, $v_2(k_\alpha) = (1 + q(k_\alpha))v_2^L - k_\alpha \tau$ and that for $k < k_2, v'(k) > 0$ so that:

$$\frac{dv_2(k_\alpha)}{dv_2^H} = \frac{\partial v_2(k_\alpha)}{\partial v_2^H} + v'(k_\alpha)\frac{dk_\alpha}{dv_2^H} = 0 + (+ \cdot -) = -$$

Putting these derivatives together with the derivative of p^m with respect to S, H, E we have the required result. □

References

James Bessen. Hold-up and patent licensing of cumulative innovations with private information. *Economics Letters*, 82(3):321–326, 2004.

James Bessen and Eric Maskin. Sequential innovation, patents, and innovation. NajEcon Working Paper Reviews 321307000000000021, www.najecon.org, May 2006.

Howard F Chang. Patent scope, antitrust policy, and cumulative innovation. *The RAND Journal of Economics*, 26(1):34–57, 1995. ISSN 07416261.

Iain Cockburn. Blurred boundaries: Tensions between open scientific resources and commercial exploitation of knowledge in biomedical research, 2005.

Victor Denicolo. Two-stage patent races and patent policy. *Rand Journal of Economics*, 31:488–501, 2000.

R. Eisenberg and M. Heller. Can patents deter innovation? the anticommons in biomedical research. *Science*, 280(5364):690–701, 5 1998.

Nancy Gallini. Patent policy and costly imitation. *Rand Journal of Economics*, 23(1): 52–63, 1992.

Jerry Green and Suzanne Scotchmer. On the division of profit between sequential innovators. *Rand Journal of Economics*, 26(1):20–33, 1995.

Bronwyn Hall and Rosemarie Ziedonis. The patent paradox revisited: an empirical study of patenting in the u.s. semiconductor industry, 1979-1995. *Rand Journal of Economics*, 32(1):101–128, 2001.

Carmen Matutues, Pierre Regibeau, and Katharine Rockett. Optimal patent design and the diffusion of innovations. *Rand Journal of Economics*, 27(1):60–83, 1996.

Peter Menell and Suzanne Scotchmer. Intellectual property, 6 2005. forthcoming, Handbook of Law and Economics. Mitch Polinsky and Steven Shavell, eds. Amsterdam: Elsevier.

Ted O'Donoghue, Suzanne Scotchmer, and Jacques Thisse. Patent breadth, patent life, and the pace of technological improvement. *Journal of Economics and Management Strategy*, 7:1–32, 1998.

Suzanne Scotchmer. Protecting early innovators: Should second-generation products be patentable? *The RAND Journal of Economics*, 27(2):322–331, 1996. ISSN 07416261.