PATENT POLICY, PATENT POOLS, AND THE ACCUMULATION OF CLAIMS IN SEQUENTIAL INNOVATION

GASTÓN LLANES[†] AND STEFANO TRENTO[‡]

Abstract. We present a dynamic model where the accumulation of patents generates an increasing number of claims on sequential innovation. We compare innovation activity under three regimes –patents, no-patents, and patent pools– and find that none of them can reach the first best. We find that the first best can be reached through a decentralized tax-subsidy mechanism, by which innovators receive a subsidy when they innovate, and are taxed with subsequent innovations. This finding implies that optimal transfers work in the exact opposite way as traditional patents. Finally, we consider patents of finite duration and determine the optimal patent length.

Keywords: Sequential Innovation, Patent Policy, Patent Pools, Anticommons, Double Marginalization, Complementary Monopoly (JEL L13, O31, O34).

Date: November 10, 2010.
[†] Escuela de Administración, Pontificia Universidad Católica de Chile, gllanes@uc.cl.
[‡] Department of Economics and Economic History, Universitat Autònoma de Barcelona, stefano.trento@uab.es.
We are grateful to Michele Boldrin for his guidance and advice. We thank Antonio Cabrales, Antonio Ciccone, Marco Celentani, Andres Erosa, Belén Jerez, Gerard Llobet, Xavier Vives, and participants of seminars at Harvard Business School, Universidad Carlos III de Madrid, Universitat Autònoma de Barcelona, and Center for Applied Economics of Universidad de Chile for useful comments and suggestions. All remaining errors are our responsibility. We gratefully acknowledge financial support from the Ministry of Education of Spain (Llanes, FPU grant AP2003-2204), the Ministry of Science and Technology of Spain (Trento, grant SEJ2006-00538), and the Comunidad Autónoma de Madrid (Trento). Stefano Trento is also affiliated to MOVE and Barcelona Graduate School of Economics.

1. Introduction

Knowledge builds upon previous knowledge. This is true for most innovations nowadays, especially in high-tech industries such as molecular biology, plant biotechnology, semiconductors, and software. In some cases, the innovation consists of an improvement of an older version of the same good. In other cases, the research leading to the discovery of the new good depends on the access to research tools, techniques and inputs that were previous innovations themselves.

The sequential nature of innovation introduces the issue of how to divide the revenues from the chain of inventions among the different innovators. Suppose two innovations may be introduced sequentially. If the first innovator receives a patent, she may obtain a claim over part of the second innovator's revenues. Then, the policy maker faces an important trade-off: if the patent covering the first innovation is strong, the second innovation may become unprofitable, but if that patent is weak, it may provide low incentives to introduce the first innovation.

The literature on sequential innovation, pioneered by Scotchmer (1991), has studied this problem in depth. Usually, this literature has analyzed the optimal division of profits between two sequential innovators. But what happens when a continuous sequence of innovations exists, each building on all previous inventions?

Recent research has suggested the possibility that the accumulation of claims on sequential innovations may generate a *tragedy of the anticommons* (Heller, 1998; Heller and Eisenberg, 1998). When too many agents have exclusion rights over the use of a common resource, this resource tends to be underutilized, in clear duality with the tragedy of the commons in which too many agents hold rights of use and the resource tends to be overused.

In our case, the anticommons could arise if too many patent holders have exclusive claims on separate components of the state-of-the-art technology, creating an obstacle for future research. However, the anticommons hypothesis has not yet been studied formally in a dynamic model with endogenous innovation. In particular, patents produce claims on subsequent innovations that may more than compensate for the negative effect of having to pay licensing fees to previous innovators. Several interesting questions arise: what is the net effect of patents on innovation incentives? How should policy parameters be set to maximize social welfare?

Two streams of literature provide partial answers to these questions.[1] The literature on sequential innovation is mainly concerned with substitute innovations, where later innovations are applications or improvements of earlier innovations. This setting allows little room for studying the effects of the accumulation of claims. On the other hand, the literatures of complementary monopoly, patent thickets, and patent pools study the problem of the accumulation of complementary patents, but from a static point of view. As Shapiro (2001) states: *"The generic problem inherent in the patent thicket is well understood as a matter of economic theory, at least in its static version."* The main contribution of our paper is to develop a dynamic model to study how the accumulation of complementary claims affects innovation incentives. We relate the literature on sequential innovation to the static literatures of complementary monopoly and patent pools. Extending the analysis of complementary monopoly to a dynamic framework allows us to gain relevant insights into the emergence of patent thickets and the net effect of different patent policy regimes on innovation at different stages of industry maturity.

We present a dynamic model to study the division of profits between sequential innovators when each innovation builds on several prior inventions. An infinite sequence of innovations $n = 1, 2, \dots$ exists, where innovation n cannot be introduced until innovation $n-1$ has been introduced. Each innovation has a commercial value (the profit it generates as a final good), which is random and private information of the innovator, and requires a deterministic cost of R&D to be developed.

Our model provides a good description of the innovation process in several industries. For example, in the software industry, the first programs were written from scratch and therefore built on little prior knowledge. As more and more programs were developed, they progressively became more dependent on technologies the first programs had introduced. According to Garfinkel et al (1991), modern software programs contain thousands of previously developed mathematical algorithms and techniques. Similar examples can be found in other high-tech industries.

Formally, our model is a multi-stage game in discrete time with an uncertain end. Interestingly, the probability of reaching the next period is determined endogenously. The equilibrium concept we use is subgame perfect equilibrium with Markovian strategies (Markov perfect equilibrium).

[1]Read section 1.1 for more details.

In the first sections of the paper, we study equilibrium dynamics in three scenarios: patents, no patents, and patent pools. With patents, innovation becomes harder and harder with more complex innovations. The probability of innovation goes to 0 as $n \to \infty$. The probability of innovation is higher than in the static case, but not high enough to prevent the tragedy of the anticommons. Therefore, we show that complementary monopoly inefficiencies, originally explored by Cournot (1838), also extend to a dynamic framework where we remove the bound on the social value complementary monopolists share, as we explain in section 1.1.

Without patents, the probability of innovation is constant and depends on the degree of appropriability of the innovation's commercial value in the final goods sector.

When patents protect ideas, the formation of a patent pool increases the probability of innovation for all innovations. Interestingly, the probability of innovation with a pool is constant and higher than it would be in the static case. This result strengthens the findings of Shapiro (2001), Lerner and Tirole (2004), and Llanes and Trento (2009) for static models.

We find that pools are dynamically unstable: the temptation to remain outside the pool increases as the sequence of innovations advances, which means early innovators have more incentives than later innovators to enter the pool. The design of a mechanism to solve the pool instability problem, along the lines of Brenner (2009), is beyond the scope of this paper. However, we find that a scheme in which each innovator buys all patent rights from the preceding innovator, instead of paying only for the permission to use the idea, can replicate the patent pool outcome. The complete sale of patent rights will therefore generate higher innovation than licensing. An alternative scheme, leading to the same innovation outcome, is to allow subsequent competition between the licensee and the original licensor. This alternative scheme removes the monopoly power of all but the last patent, eliminating the anticommons effect.

We study the optimal innovation policy that maximizes the expected welfare of the sequence of innovations and find that innovation is suboptimal in the three policy regimes. In the no-patents regime, there is a dynamic externality: innovators do not consider how their decisions impact the technological possibilities of future innovators. In the two other policy regimes, the inefficiency stems from asymmetric information and market power: patent holders do not know the exact value of the innovation, but they know its probability distribution. The asymmetric information generates a downward-sloping expected demand for the use of ideas, and patent holders' market power results in a price for old ideas above their marginal cost.

We also show that the first best can be reached by decentralizing the innovation decision and implementing a tax-subsidy scheme. Surprisingly, this result holds even if the government and previous innovators do not know the value of the innovation. With respect to the timing of the optimal transfers, we find that innovators should receive a subsidy to innovate, and then be taxed with subsequent innovations. Therefore, optimal transfers work in the exact opposite way as patents, which require an innovator to pay previous inventors upfront and then be compensated by following innovators.

These findings extend the results of Erkal and Scotchmer (2007), who show that the scarcity of ideas has a non-trivial impact on the optimal reward to the innovator. In this sense, the two papers are complementary. While Erkal and Scotchmer show that the scarcity of ideas affects the *size* of the optimal reward, we show that it also affects the *timing* of this reward. Finally, we find that symmetric information over the value of the innovation leads to the first best. We then turn to the analysis of the optimal patent length.

With respect to the optimal patent length, we find that short patents maximize the probability of innovation, because in our model, absent any form of price collusion or agreement between patent holders, reducing patent length is the only way to reduce the complementary monopoly problem. This finding complements the findings of previous papers that focus on the final goods sector and substitute innovations.

Another important finding is that the main results of the literature of complementary monopoly extend to a dynamic framework with endogenous innovation. Also, patent pools composed of complementary patents are welfare improving with endogenous innovation. This finding partially answers a question Lerner and Tirole (2004) posed regarding the desirability of patent pools when their *ex-ante* effect on innovation activity is taken into account.

1.1. **Related Literature.** We extend the literature on sequential innovation by analyzing the case in which patents generate cumulative claims on subsequent innovations. This extension is important because it allows us to study the emergence of patent thickets. Many papers of sequential innovations, such as Green and Scotchmer (1995), Chang (1995), and Scotchmer (1996), analyze the optimal distribution of profits between two sequential innovators. If the first innovation has low commercial value (basic research for instance), granting the first innovator a strong patent is optimal. This is not necessarily true in our model, where the accumulation of patents generates a problem of anticommons.

Abstracting from transaction costs, or the possibility that one or more patent holders refuse to license their ideas thereby blocking innovation, the tragedy of the anticommons is similar to a complementary monopoly problem, first analyzed by Cournot (1838). Cournot modeled a competitive producer of brass who has to use copper and zinc as inputs in production, and showed that, when two different monopolists sell the inputs, the total cost of producing brass is higher than when the same monopolist sells both inputs. Sonnenschein (1968) showed that complementary monopoly is the dual of a duopoly model with quantity competition and homogeneous goods, and Bergstrom (1978) generalized this result to a general number n of inputs and any degree of complementarity among them. Chari and Jones (2000) showed that the market outcome in a complementary monopoly setting is increasingly inefficient as the number of agents increases.

Recently Shapiro (2001) and Lerner and Tirole (2004) applied the instruments of complementary monopoly to the analysis of patent pools. Their results reinforce the results on complementary monopoly: patent pools (or, equivalently, a single monopolist owning all the patented inputs) reduce the cost of innovation when patents are complements and increase the cost when patents are substitutes. Boldrin and Levine (2005) and Llanes and Trento (2009) also made use of complementary monopoly to show that, as the number of complementary patents increases, the probability that a future innovation will be profitable goes to zero.

All of these papers, although they make important contributions, present static models. In other words, the first innovation has been invented already, so patents and patent pools only affect the profitability of introducing a second innovation. This structure introduces an important asymmetry between previous and future innovations that our dynamic model eliminates. We believe that adding a dynamic dimension is an important step towards a better understanding of the mechanism of anticommons in sequential innovation.

In particular, one would expect the complementary monopoly problem to be weaker in a dynamic context for two reasons: first, in the static model there is a limit on the revenues input producers share; we eliminate this limit by extending the analysis to a dynamic framework with potentially infinite innovations. Second, setting high license fees increases the probability that the innovation chain, or a particular research line, will come to a halt: a patent holder would then have an incentive to moderate the license fee to be able to reap part of the revenues of further innovations. We find that, in spite of these two effects, the complementary monopoly problem is so strong that innovation eventually becomes unprofitable.

Our paper is related to O'Donoghue et al (1998) and Hopenhayn et al (2006), who also present models of cumulative innovation. However, Hopenhayn et al (2006) have no accumulation of claims since only one patent is valid at any given time. O'Donoghue et al (1998), on the other hand, have accumulation of claims but rule out complementary monopoly, as bargaining among patent holders is efficient by assumption. Also, in both papers, innovations are substitutes: the introduction of a new product automatically implies the disappearance of old versions from the market. The substitutability between innovation introduces a natural limit on appropriability and produces an important trade-off, because granting a patent to the first innovator limits what can be offered to the second innovator. In our paper, innovations are complementary and do not compete with each other in the final goods sector. This setting eliminates the appropriability problem. In such a model, one would expect a patent system to perform well. However, the opposite happens: granting too many patent rights on sequential innovations produces a complementary monopoly problem that hampers innovation.

Finally, our paper is also related to Menezes and Pitchford (2004), who present a dynamic model of anticommons. Menezes and Pitchford model the case of a buyer who has to combine complementary assets from two sellers. Sellers may have an incentive to avoid entering into negotiations with the buyer because they may get a higher share of total surplus by negotiating after the buyer agrees with the other seller. Holdout occurs if at least one seller is not present in the first round of negotiations. The authors show that complementarity is a necessary condition for holdout, and also that a rise in complementarity leads to an increase in the possibility of holdout. In our case, the accumulation of claims may lead to increasing delays in the agreement between current and past innovators, furthering the welfare loss caused by complementary monopoly.

2. The model

We study a model with an infinite sequence of innovations $n = 1, 2, \ldots$ Each innovation cannot be introduced until all previous innovations have been introduced. This innovation process reflects the fact that earlier innovations do not have a solid background upon which to build, while further innovations become more and more indebted to previous ones as the market matures.

At each stage, an innovator gets an idea of how to develop a particular innovation. If the innovator decides to perform the innovation, the game continues and, in the following stage, another innovator will get an idea for the next innovation. If the innovator decides not to introduce the innovation, two things may happen: (i) with probability ϕ, the game continues and in the following stage another innovator tries to perform the failed innovation, and (ii) with probability $1-\phi$, the game ends and no other innovations are possible.

The parameter $\phi \in [0,1]$ represents the degree of scarcity of ideas. If ideas are more scarce (lower ϕ), each idea is more difficult to substitute, and another innovator is less likely to have a different approach to implement a failed innovation.

Let n, j represent the j^{th} innovator trying to introduce innovation n ($j-1$ innovators have already tried to introduce innovation n without success). At the beginning of the stage, the innovator gets an idea with *random* value $v_{n,j}$, which she may develop by incurring in a *deterministic* R&D cost of ε.

$v_{n,j}$ represents the revenues obtained by selling the new product in the final-goods market. To concentrate on the effects of patents on innovation activity, we will assume the innovator is a perfect price discriminator in the final-goods market, which means the private value of the innovation is equal to the social surplus the new product generates.

The value of the innovation is private information of the innovator. Patent holders only know $v_{n,j}$ is drawn from a uniform distribution between 0 and 1, with cumulative distribution function $F(v_{n,j}) = v_{n,j}$.

The innovator's decision on whether to perform the innovation will depend not only on $v_{n,j}$ and ε, but also on the licensing revenues and cost that may arise depending on the particular patent regime under analysis.

Given that at each stage the innovator will perform the innovation with a certain probability, the game is a multi-stage game with uncertain end, in which the probability that the game continues is determined endogenously.

3. Innovation with patents

In this case, patents with infinite length and breadth protect ideas (we will relax these assumptions in section 10), which means each innovator has to pay license fees to all previous inventors (patent holders), in case she wants to introduce the innovation. The cost of innovation is the sum of the cost of R&D and the licensing fees paid to previous innovators.

A patent of infinite length will also protect the new idea, which means the innovator can request licensing fees from all subsequent innovators. The total revenues of the innovation equal the commercial value of the innovation plus future licensing revenues.

The timing of the game within each stage is the following: (i) The $n-1$ patent holders set licensing fees $p^i_{n,j}$, (ii) Nature extracts a value for $v_{n,j}$ from distribution $F(v_{n,j})$, (iii) the innovator decides whether to innovate ($I_{n,j}=1$) or not ($I_{n,j}=0$).

If the revenues from the innovation are higher than the cost, innovator n,j will introduce the innovation, and in the next stage, innovator $n+1,1$ will try to introduce innovation $n+1$. If revenues are lower than cost, innovator n,j will not introduce the innovation, and in the following stage (reached with probability ϕ), innovator $n,j+1$ will try to introduce innovation n based on a different approach. This innovator $n,j+1$ will face the same $n-1$ patent holders and will have a new draw for the value of innovation, $v_{n,j+1}$.

Let $J^i_{n,j}$ be the expected future licensing revenues of patent holder i at trial j of innovation n, given that stage n,j has been reached. Expressed in a recursive way,

$$J^i_{n,j} = Pr_{n,j}\,(p^i_{n,j} + \beta\, J^i_{n+1,1}) + (1 - Pr_{n,j})\,\phi\,\beta\, J^i_{n,j+1},$$

where $Pr_{n,j}$ is the probability that innovation n is introduced at trial j, given that $n-1$ prior innovations have been introduced and that $j-1$ trials to introduce innovation n have already failed. With probability $Pr_{n,j}$, the patent holder gets the price $p^i_{n,j}$ plus the continuation value of the first trial of the next innovation, $J^i_{n+1,1}$, discounted by a factor $\beta \in [0,1]$. With probability $(1 - Pr_{n,j})\,\phi$, the innovation is not introduced but the game continues, in which case the patent holder gets the continuation value corresponding to the next trial of the current innovation, $J^i_{n,j+1}$, discounted by the factor β. β can be interpreted both as the discount factor or, for a fixed discount factor, as the time between innovations: lower values of β imply ideas arrive less frequently.

The innovator's payoff is $I_{n,j}(v_{n,j} + \beta J^n_{n+1,1} - c_{n,j} - \varepsilon)$, where $c_{n,j} = \sum_{i=1}^{n-1} p^i_{n,j}$ is the sum of licensing fees paid to previous innovators.

We will focus on Markov strategies. A strategy for player i specifies an action conditioned on the state, where actions are prices and the state is simply n,j. The equilibrium concept is Markov perfect equilibrium, which implies future prices will be determined by a Nash equilibrium in the subsequent games. Thus players understand that no action taken today can influence future prices and probabilities. Current actions probabilistically affect state transitions through the influence of current prices on the probability of innovation. We have just proved the following lemma:

Lemma 1. *$J^i_{m,k}$ does not depend on any action taken at stage n, j for $m > n$ and any j, k.*

The game is solved recursively. The solution to the innovator's problem is straightforward. Given $v_{n,j}$ and $c_{n,j}$, the innovator forecasts $J^n_{n+1,1}$, and decides to innovate ($I_{n,j} = 1$) if the revenues from the innovation exceed the cost of innovation:

$$I_{n,j} = \begin{cases} 1 & \text{if } \ v_{n,j} + \beta J^n_{n+1,1} \geq c_{n,j} + \varepsilon, \\ 0 & \text{otherwise,} \end{cases}$$

which implies that the probability of innovation is $Pr_{n,j} = 1 + \beta J^n_{n+1,1} - c_{n,j} - \varepsilon$.

At stage n, j, patent holders want to maximize their expected licensing revenues from stage n, j onwards. They know their decisions do not affect $J^i_{m,k}$ for any k, j and $m > n$ (they can only affect the probability that stage m, k is reached) and decide a licensing fee $p^i_{n,j}$, taking the decisions of the other patent holders as given. The patent holder's problem is

$$\max_{p^i_{n,j}} J^i_{n,j} = Pr_{n,j}\,(p^i_{n,j} + J^i_{n+1,1}) + \phi\,\beta\,(1 - Pr_{n,j}) J^i_{n,j+1}.$$

From the first-order conditions we obtain that the optimal price is $p^i_{n,j} = \frac{1-\varepsilon+\phi\beta J^i_{n,j+1}}{n}$. Bellman equations are easy to solve because Lemma 1 implies current patent holders take future prices and innovation decisions as given when setting their licensing fees.

Imposing symmetry, $p^i_{n,j} = p_{n,j}$ and $J^i_{n,j} = J_{n,j}$ for all i. Also, the problem at trial j is the same as the problem at trial k for any j, k, which means $J_{n,j} = J_{n,k} = J_n$. Substituting the probability of innovation, we get $Pr_n = \frac{1-\varepsilon}{n} + \beta J_{n+1} - \frac{n-1}{n}\phi\beta J_n$, and substituting this result into the expression for $J^i_{n,j}$:

$$J_n = \frac{1}{1-\phi\beta}\left(\frac{1-\varepsilon}{n} + \beta J_{n+1} - \frac{n-1}{n}\phi\beta J_n\right)^2.$$

Rearranging this expression,

$$J_{n+1} = \frac{1}{\beta}\left(\sqrt{(1-\phi\beta)J_n} - \frac{1-\varepsilon}{n}\right) + \frac{n-1}{n}\phi J_n,$$

which is a decreasing sequence converging to 0 as $n \to \infty$.

The sequence in terms of probabilities is:

$$Pr_{n+1}^2 = \frac{1-\phi\beta}{\beta}\left(Pr_n - \frac{1-\varepsilon}{n}\right) + \frac{n-1}{n}\phi Pr_n^2,$$

which is also a decreasing sequence converging to 0 as $n \to \infty$. The result is therefore that innovation gets harder and harder with more complex innovations (those that build upon a larger number of previous innovations).

The probability of innovation decreases with complexity because patent holders do not take into account cross-price effects: patent holder i sets the price of her patent by equating the marginal revenue and the marginal cost of increasing her license fee. The marginal revenue is simply the additional revenue in case the new innovation is performed. The marginal cost is the reduction in expected demand, and depends on the fact that – since all patents are essential for the new innovation – increasing the price of patent i decreases the probability of innovation. Increasing the price of patent i will also reduce the expected demand for all other inputs. But patent holder i will not take this effect into account, generating the anticommons effect that closely resembles the tragedy of the commons: patent holders ignore cross-price effects and set prices that are higher than they would set if they were coordinated (see section 5).

4. Innovation without patents

Suppose a policy reform completely removes patents. This change affects innovation in two ways. First, the revenues of the innovator in the final-goods sector will decrease as a result of imitation. Specifically, assume that the innovator can only appropriate a fraction $\theta \in [0,1]$ of the consumer surplus the innovation generates. Second, innovators will not pay licensing fees to previous innovators, nor will they charge for the use of their ideas in subsequent innovations. Therefore, $c_{n,j} = 0$ and $J_{n,j} = 0$ in the previous model.

The timing of the game is the following: (i) nature extracts a value of the innovation $v_{n,j}$, and (ii) the innovator decides whether to innovate.

The innovator will innovate if $\theta\, v_{n,j} \geq \varepsilon$ and will not innovate otherwise. Thus, the probability of innovation is constant and equal to $1 - \varepsilon/\theta$ if $\theta > \varepsilon$. If $\theta \leq \varepsilon$, then the probability of innovation is zero.

5. Patent pools

In this section, we analyze what happens when a collective institution such as a patent pool sets licensing fees cooperatively. At each stage, the pool maximizes the future expected revenues of *current* patent holders. The pool will set a symmetric price for all current patent holders. Once an innovation is performed, the innovator becomes a member of the pool in all subsequent stages. The first stage has no pool because no innovation has been introduced (the pool plays from stage 2 onwards).

The probability of innovation is $Pr_{n,j} = 1 + \beta J^n_{n+1,1} - (n-1)p_{n,j} - \varepsilon$, and the pool's problem is

$$\max_{p_{n,j}} J^i_{n,j} = Pr_{n,j}\,(p_{n,j} + \beta J^i_{n+1,1}) + (1 - Pr_{n,j})\,\phi\,\beta\,J^i_{n,j+1}.$$

The difference with respect to the non-cooperative case is that the pool recognizes cross-price effects, and therefore is encouraged to set lower prices than in the no-pool case.

By symmetry, $J^i_{n+1} = J_{n+1}$. A higher J_{n+1} fosters innovation in two ways. First, it increases the innovator's future revenues. Second, it encourages the pool to set a lower price, because it increases the loss of current patent holders if the sequence of innovations is stopped.

The equilibrium price is $p_{n,j} = \frac{1-\varepsilon}{2(n-1)} - \frac{n-2}{2(n-1)}\beta J^i_{n+1,1} + \frac{1}{2}\phi\beta J^i_{n,j+1}$, which is equal to the price a pool would set in a static model (see section 9.3) *minus* an additional term arising from the pool's concern for keeping future revenues.

The probability of innovation becomes $Pr_n = \frac{1-\varepsilon}{2} + \frac{n}{2}\beta J_{n+1} - \frac{n-1}{2}\phi\beta J_n$. Introducing price and probability in J_n, we get

$$J_n = \frac{1}{\phi\beta(n-1)}\left(\frac{1-\varepsilon}{2} + \frac{n}{2}\beta J_{n+1} + \frac{n-1}{2}\phi\beta J_n\right)^2.$$

Rearranging this expression,

$$J_{n+1} = \frac{1}{\beta n}\left(\sqrt{(1-\phi\beta)(n-1)J_n} - (1-\varepsilon)\right) + \frac{n-1}{n}\phi J_n,$$

which is a decreasing sequence converging to 0 as $n \to \infty$.

The sequence in terms of probabilities is

$$Pr^2_{n+1} = \frac{2(1-\phi\beta)}{\beta}\left(Pr_n - \frac{1-\varepsilon}{2}\right) + \phi Pr^2_n,$$

which is a constant sequence such that

$$Pr_n = \frac{1 - \phi\beta - \sqrt{1-\phi\beta}\sqrt{1-\beta+\beta(1-\phi)\varepsilon}}{\beta(1-\phi)}$$

for $n \geq 2$ and $Pr_1 = \min\{1, \frac{2-2\sqrt{1-\beta(1+\varepsilon(-1+\phi))}\sqrt{1-\beta\phi}+\beta\phi(-3+\varepsilon+\phi(1-\varepsilon))}{\beta(1-\phi)^2}\}$.

6. Comparison

Figure 1 shows the evolution of the probability of innovation in the three cases studied above: infinitely lived patents, no-patents and patent pool. In the figure, $\beta = 0.95$, $\phi = 0.5$, and the cost of R&D is $\varepsilon = 0.2$, but the qualitative results discussed in this section hold for any choice of these parameters. We consider $\theta = 1$ (full appropriation) and $\theta = 0.3$ (the innovator appropriates 30% of the social surplus generated by the new product) for the no-patents case.

Comparing the patent and no-patents cases, we can see that patents increase the probability of the first innovations but decrease the probability of further innovations. The number of innovations for which patents increase the probability depends on θ. For example, when $\theta = 1$, patents only increase the probability of the first innovation. Nevertheless, even when $\theta = 0.3$, the probability increases only for the first two innovations. For patents to increase the probability of several innovations, θ must be very small and close to ε (i.e., when very little appropriability exists without patents).

When ideas are protected by patents, the formation of a patent pool increases the probability of innovation. Figure 1 shows that, when patents protect ideas, the probability of innovation is always larger with patent pools than without it. Moreover, with a pool, the probability of innovation does not go to zero as $n \to \infty$. The comparison between patent pools and no-patents depends on ε and θ. When θ is low, a patent pool increases the probability of all innovations. When θ is high, the pool increases the probability of the first innovation, and decreases the probability of all subsequent innovations.

7. Complete sale of patent rights

The tragedy of the anticommons stems from fragmented ownership of complementary patents. In this case, the probability of innovation decreases as more innovations are introduced, converging to 0 as $n \to \infty$. The formation of a patent pool would alleviate this problem by concentrating all pricing decisions on one entity. In this section, we discuss a possible alternative solution, which is to enforce the sale of complete patent rights instead of allowing the sale of individual access rights through licensing fees. Other innovators can, in turn, purchase these patent rights. In this case, innovator 1 would sell the complete patent rights over innovation 1 to innovator 2 for a price r_1. Innovator 2 then would sell the patent rights on innovations 1 and 2 to innovator 3 for a price r_2, and so on. We will show that this mechanism eliminates

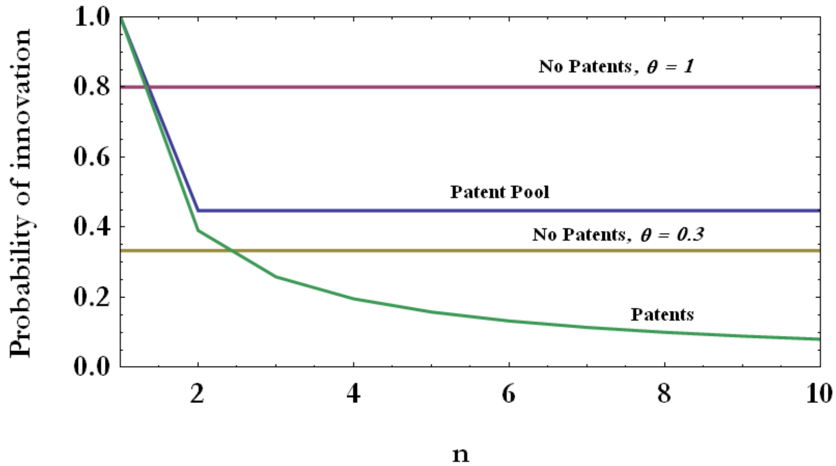


FIGURE 1. Comparison of equilibria.

the coordination failure at the basis of the anticommons, and that it replicates the innovation outcome under a patent pool.

The cost of innovation n, j becomes $\varepsilon + r_{n-1,j}$, and expected revenues $v_{n,j} + \beta J^s_{n+1,1}$, where $J^s_{n+1,1}$ are the expected revenues of innovator n from selling the n patent rights to innovator $n+1, 1$.

The probability that innovation n is performed is $Pr_{n,j} = 1 - \varepsilon - r_{n-1,j} + \beta J^s_{n+1,1}$. At stage n, j, innovator $n-1$ solves the following maximization problem:

$$\max_{r_{n-1,j}} J^s_{n,j} = Pr_{n,j}\, r_{n-1,j} + \phi\beta(1 - Pr_{n,j}) J^s_{n,j+1}.$$

Solving the maximization problem, and given that the problem is the same for any j, yields a price for patent rights $r_{n-1} = \frac{1-\varepsilon+\beta J^s_{n+1}+\beta\phi J_n}{2}$. The resulting sequence of probabilities of innovation is:

$$Pr^2_{n+1} = \frac{2(1-\phi\beta)}{\beta}\left(Pr_n - \frac{1-\varepsilon}{2}\right) + \phi Pr^2_n,$$

which is exactly the same sequence as with patent pools.

We have just proved that the complete sale of patent rights is equivalent to a patent pool in our dynamic model. However, note that implementing this scheme may be difficult when describing the nature of innovations ex-ante is hard. For example, when selling the rights over innovation n to innovator $n+1$, describing what innovation $n+2$ may be is difficult. In this case, complete contracts may be hard to write, making patent pools easier to enforce.

An alternative policy arrangement leading to the same result would be the following: restoring the possibility of licensing access rights, but at the same time allowing subsequent competition between the licensee and the original licensor. In this case, if innovator n licenses the use of innovation n to innovator $n+1$, then innovator $n+2$ can license the use of innovation n from innovators n and $n+1$. Under this policy arrangement, innovator n will only get positive revenues from licensing her innovation to innovator $n+1$ because, at stage $n+1$, she is a monopolist. After stage $n+1$, she will face competition from other innovators, and Bertrand competition will imply a licensing fee equal to zero.

8. Endogenous patent pool formation

In section 5 we assumed all innovators, after innovating, automatically join the patent pool. In this section, we endogeneize this choice, by analyzing the incentives for innovator $n-1$ to join the pool. In particular, we compare the expected revenues from joining the pool ($J_{n,j}$ from section 5) with the expected revenues from setting the price of her patent non-cooperatively ($J^O_{n,j}$).

We start with the non-cooperative choice. For expositional clarity, let us refer to the patent pool members as *insiders* and to the non-cooperative member as the *outsider*. The pool maximizes the expected revenues of the insiders:

$$\max_{p^i_n} J^I_{n,j} = Pr_{n,j}\,(p^I_{n,j} + \beta J^I_{n+1,1}) + (1 - Pr_{n,j})\,\phi\beta J^I_{n,j+1},$$

where $p^i_{n,j}$ is the cooperative price of an insider's patent, and $Pr_n = 1 + J^I_{n+1,1} - (n-2)p^I_{n,j} - p^O_{n,j} - \varepsilon$, with $p^O_{n,j}$ denoting the price of the outsider's patent.

On the other hand, the outsider maximizes

$$\max_{p^O_n} J^O_{n,j} = Pr_{n,j}\,(p^O_{n,j} + \beta J^O_{n+1,1}) + (1 - Pr_{n,j})\,\phi\beta J^O_{n,j+1}.$$

From first-order conditions we know that $J^O_{n,j} = (n-2)\,J^I_{n,j}$, meaning that, if there is an outsider in equilibrium, she will have higher profits than each of the insiders.

Now let us compare the expected revenues from not joining the pool ($J^O_{n,j}$) with the expected revenues of joining the pool, given that everybody else is in the pool ($J_{n,j}$ from section 5). In equilibrium, deviating from the pool produces and expected revenue of

$$J^O_n = \frac{3(1-\phi\beta) - \sqrt{1-\phi\beta}\sqrt{9 - 8(1-\varepsilon(1-\phi)) + \phi}}{16\beta^2(1-\phi)^2(1-\phi\beta)},$$

which does not depend on n. If, on the other hand, innovator $n-1$ decides to join the pool with the $n-2$ previous innovators, her expected revenue will be

$$J_n = \frac{\left(1 - \phi\beta - \sqrt{1-\phi\beta}\sqrt{1-\beta+\beta(1-\phi)\varepsilon}\right)^2}{(n-1)(1-\phi\beta)\beta^2(1-\phi)^2},$$

which is decreasing in n. This result is due to the fact that the patent pool maximizes joint profits, thus keeping the total cost of innovation constant. This constant amount must be divided among an increasing number of insiders; therefore, the expected revenue of an insider is decreasing in n and converges to 0 as $n \to \infty$. Therefore some innovator $n_0 > 2$ with incentives to deviate by remaining outside the pool always exists.

Figure 2 illustrates this finding and shows the gains from deviating from the pool as a function of n, for $\varepsilon = 0.1$, $\phi = 0.5$, and $\beta = 0.95$. The gains become positive after innovator 3, which means the fourth innovator would gain by remaining outside the pool.

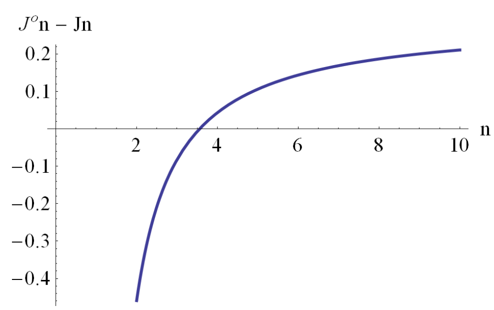


FIGURE 2. Gains from not joining the patent pool. $\varepsilon = 0.1$, $\phi = 0.5$, and $\beta = 0.95$

Patent pools can improve innovation activity, but are dynamically unstable. Early innovators have more incentives to enter the pool than subsequent innovators. Brenner (2009) finds an elegant mechanism to solve the instability problem for socially desirable patent pools in a static model. We leave the design of an equivalent mechanism in the context of a dynamic model for future research. Without such a mechanism, however, patent pools are likely to be unstable. This instability might explain why governments sometimes have to enforce the creation of patent pools, as the U.S. government did in the radio and aircraft industry, for example.

9. Socially optimal innovation

The relevant measure of welfare is the expected social value generated by the sequence of innovations. The social value of an innovation is equal to the increase in consumer surplus minus the cost of the resources spent in R&D. Therefore, when considering trial j of innovation n, the social value generated is $v_{n,j} - \varepsilon$ if the innovation is performed, and 0 otherwise.

Consider the decision of performing innovation n, j. If the value of the innovation is greater than the cost, obviously the innovation should be performed. However, the social planner could still decide to perform an innovation with negative social value, because in the opposite case, the sequence of innovations will stop with a probability of $1 - \phi$. The decision will depend, therefore, on a comparison between the current cost of performing an innovation with negative social value and the expected future benefits of continuing with the chain of innovations.

Let $W_{m,k}$ be the expected social welfare from stage m, k onwards. Once we know the realization of $v_{n,j}$, expected welfare is $v_{n,j} - \varepsilon + \beta\, W_{n+1,j}$ if the innovation is performed, and $\beta\,\phi\, W_{n,j+1}$ if the innovation is not performed. Therefore, the innovation should be performed if $v_{n,j} - \varepsilon + \beta\, W_{n+1,j} \geq \beta\,\phi\, W_{n,j+1}$.

Proposition 1 shows the socially optimal innovation policy.

Proposition 1 (Socially optimal innovation)**.** *To maximize expected social welfare, innovation n, j should be performed if and only if $v_{n,j} \geq \underline{v}^*$, where*

$$\underline{v}^* = \begin{cases} 0 & \text{if } \varepsilon \leq \frac{\beta}{2}\,\frac{1-\phi}{1-\beta\,\phi}, \\ \frac{\beta-1+\sqrt{1-\beta\,\phi}\,\sqrt{1-2\,\beta\,(1-(1-\phi)\,\varepsilon-\phi/2)}}{\beta\,(1-\phi)} & \text{if } \varepsilon > \frac{\beta}{2}\,\frac{1-\phi}{1-\beta\,\phi}. \end{cases}$$

Proof. Given the assumptions of the model, innovations $v_{n,j}$ and $v_{m,k}$ are equivalent for any n, j, m, k. It follows that $W_{n,j} = W_{m,k} = W$, and the optimal decision is time-invariant: a value $\underline{v}^* \in [0, \varepsilon]$ exists such that innovation n, j should be performed if and only if $v_{n,j} \geq \underline{v}^*$. By definition, $\underline{v}^*$ solves $\underline{v}^* - \varepsilon + \beta\, W = \beta\,\phi\, W$. Therefore, we need to determine the value of W. In particular, $W_{m,k}$ is given by

$$\begin{aligned} W_{m,k} \;=\;& Pr(v_{m,k} \geq \underline{v}^*)\,(E(v_{m,k} - \varepsilon / v_{n,j} \geq \underline{v}^*) + \beta\, W_{m+1,k}) + \\ & (1 - Pr(v_{m,k} \geq \underline{v}^*))\,\beta\,\phi\, W_{m,k+1}. \end{aligned}$$

Imposing $W_{m,k} = W_{m+1,k} = W_{m,k+1} = W$, and solving for W, we get

$$W = \frac{1 - \underline{v}^*}{1 - \beta\,(1 - (1-\phi)\,\underline{v}^*)}\left(\frac{1 + \underline{v}^*}{2} - \varepsilon\right).$$

Substituting this result into $\underline{v}^* - \varepsilon + \beta\, W = \beta\, \phi\, W$, and solving for $\underline{v}^*$, we get the optimal policy stated in the proposition. ■

Proposition 1 implies that innovation will be suboptimal in the three cases studied above. There are three reasons why this is so: dynamic externalities, market power, and asymmetric information.

The dynamic externality is best described by analyzing the no-patents case. Without patents, the innovator will perform the innovation when $v_n \geq \varepsilon/\theta$. Given that $\underline{v}^* \leq \varepsilon$, the innovator may decide not to perform a socially desirable innovation, even if $\theta = 1$, because she ignores the effect of her decision on the technological possibilities of future innovators. This effect is well known in the literature of sequential innovation (Scotchmer, 1991; Hopenhayn et al, 2006) and is similar to the one found in the literature of moral hazard in teams (e.g., Holmstrom, 1982), where each agent internalizes only his reward from the effort exerted.

The solution to the first problem would require intertemporal transfers. In section 9.1, we show that the first best can be reached by decentralizing the innovation decision and implementing a tax-subsidy scheme. Surprisingly, this result holds even if information is asymmetric, that is, if neither the government nor previous innovators know the value of the innovation.

In the patents and patent-pool cases, the inefficiency arises from a different source: market power and asymmetric information. Because patent holders care about the stream of future licensing revenues they will lose if the sequence of innovations stops, they internalize the dynamic externality. However, asymmetric information implies a downward-sloping expected demand for old innovations, and market power implies inefficient pricing of patents, which leads to suboptimal innovation. As the number of holders of rights on innovation increases, the inefficiency due to market power increases (because of the complementary monopoly), which is why the patent-pools case is more efficient than the patents case.

In order to show the importance of the asymmetric information assumption, in Section 9.2, we show that under symmetric information there exists an equilibrium that reaches the first best.[2] This means that without asymmetric information, the dynamic externality could be perfectly internalized, reaching the first best.

[2]A continuum of equilibria exists under symmetric information. Some of these equilibria do not reach the first best, but the important fact is that some of them do.

9.1. **Optimal transfers.** In this section, we show that the first best can be reached by decentralizing the innovation decision and implementing a tax-subsidy scheme. This is a surprising finding, because it does not require the government to know the value of the innovation in order to be implemented. In addition, we find that the optimal timing of the tax-subsidy scheme is as follows: innovators should receive a subsidy if they innovate, and then be taxed when the following innovation is performed.

The structure of transfers is the following: if innovator n, j decides to innovate, she will have to pay a transfer t_n to the innovator who successfully performed innovation $n-1$, but she will also have the right to receive a transfer t_{n+1} from the innovator who performs innovation $n+1$. Therefore, given $v_{n,j}$, the innovator will innovate if $v_{n,j} - \varepsilon - t_n + \beta\, J_{n+1,1} \geq 0$, where

$$J_{n+1,1} = Pr_{n+1,1}\, t_{n+1} + (1 - Pr_{n+1,1})\, \beta\, \phi\, J_{n+1,2}. \tag{1}$$

Proposition 2 shows the optimal intertemporal transfer that implements the first best.

Proposition 2. *The optimal transfer is constant and equal to*

$$t^* = \frac{(\underline{v}^* - \varepsilon)\,(1 - \phi\,\beta\,\underline{v}^*)}{1 - \beta\,(1 - \underline{v}^* + \phi\,\underline{v}^*)}.$$

$t^* \leq 0$ *for any value of the parameters, and* $t^* < 0$ *if and only if* $\varepsilon > 0$, $\phi\,\beta < 1$.

Proof. The problems of innovators n, j and m, k are equivalent for the social planner for any n, j, m, k, which means $t_n = t_{n+1} = t$, and $J_{n,j} = J_{m,k} = J$ for any n, j, m, k. Given transfers, the probability of innovation is $Pr = 1 - \varepsilon - t + \beta\, J$. We want to make this probability equal to the optimal probability, which is $Pr^* = 1 - \underline{v}^*$. The optimal transfer then solves $\underline{v}^* = \varepsilon + t - \beta\, J$. On the other hand, from equation (1), we get $J = (1 - \underline{v}^*)\, t / (1 - \phi\,\beta\,\underline{v}^*)$. Substituting the latter expression into the former, and solving for t, we get the optimal transfer stated in the proposition. Finally, $t^* \leq 0$ because $\underline{v}^* \leq \varepsilon$ from Proposition 1. ∎

Note that $\underline{v}^*$ has a kink when $\varepsilon = \frac{\beta}{2}\,\frac{1-\phi}{1-\beta\,\phi}$, so t^* will also have a kink at that point. An interesting feature of the optimal transfer is that it is negative. Therefore, the innovator should receive a subsidy to innovate, and be taxed with the following innovation. Most importantly, optimal transfers work in the exact opposite way as patents, which require an innovator to pay previous inventors upfront and then be compensated by following innovators.

9.2. **Symmetric information.** To analyze the reasons for inefficiency in the different cases, in this section we study the effects of removing the asymmetric information assumption while keeping the market power assumption (i.e., past innovators are price-setters, whereas the current innovator is a price taker).

At each stage, a value of $v_{n,j}$ is drawn from $F(v_{n,j})$, and previous innovators set the levels of licensing fees the innovator will pay, just as in the basic model. The difference is that now, previous innovators learn the realization of $v_{n,j}$, and use this information when setting their licensing fees.

In equilibrium, previous innovators will set a level of fees that will leave the innovator indifferent between innovating or not. Otherwise, one of the previous innovators could raise her fee without affecting the innovation decision, thereby raising her profits. Therefore, any sequence of prices such that

$$\sum_{i=1}^{n-1} p_{n,j}^{i} = v_{n,j} - \varepsilon + \beta\, J_{n+1,1}^{n}$$

is an equilibrium.

Consider an equilibrium in which each innovator pays a fee only to the previous innovator:

$$p_{n,j}^{n-1} = v_{n,j} - \varepsilon + \beta\, J_{n+1,1}^{n}.$$

For the remainder of this section, let $p_{n,j} = p_{n,j}^{n-1}$ and $J_{n,j} = J_{n,j}^{n-1}$. If $v_{n,j} - \varepsilon + \beta\, J_{n+1,1} < 0$, innovator $n-1$ will not allow innovator n, j to innovate (she can do this by setting any price above $v_{n,j} - \varepsilon + \beta\, J_{n+1,1}$). In case $v_{n,j} - \varepsilon + \beta\, J_{n+1,1} \geq 0$, on the other hand, innovator $n-1$ will allow innovation n, j only if $p_{n,j} \geq \beta\, \phi\, J_{n,j+1}$. Proposition 3 shows that in equilibrium, innovations will be performed only if they are socially desirable.

Proposition 3 (Symmetric information)**.** *Under symmetric information, the equilibrium in which each innovator only pays a licensing fee to the previous innovator is socially optimal.*

Proof. The problem at trial j is the same as the problem at trial $j+1$, which means the lowest value of $v_{n,j}$ that innovator $n-1$ will tolerate is constant. Let $\underline{\hat{v}}$ indicate this value. $\underline{\hat{v}}$ solves $\underline{\hat{v}} - \varepsilon + \beta\, J_{n+1,1} = \beta\, \phi\, J_{n,j+1}$, and $J_{n,j+1}$ solves

$$J_{n,j+1} = Pr\, E(p_{n,j+1}/v_{n,j+1} \geq \underline{\hat{v}}) + (1 - Pr)\, \beta\, \phi\, J_{n,j+2},$$

where $Pr = Pr(v_{n,j+1} \geq \underline{\hat{v}}) = 1 - \underline{\hat{v}}$, and $E(p_{n,j+1}/v_{n,j+1} \geq \underline{\hat{v}}) = \frac{1+\hat{v}}{2} - \varepsilon + J_{n+1,1}$. The problem for a different n and/or j is equivalent, so $J_{n,j} = J_{m,k} = J$ for all n, j, m, k. Using this result in the above equations, we obtain

$$\begin{aligned} \underline{\hat{v}} &= \varepsilon - \beta(1 - \phi\,\beta)\,J \\ J &= (1 - \underline{\hat{v}}) \left(\frac{1 + \underline{\hat{v}}}{2} + \beta\,\phi\,J \right) + \underline{\hat{v}}\,\phi\,\beta\,J. \end{aligned}$$

Solving this system of equations for $\underline{\hat{v}}$ and J, we get that $\underline{\hat{v}} = \underline{v}^*$. ■

9.3. **Static versus dynamic incentives.** Previous models of complementary monopoly, sequential innovation, and patent pools were static (Shapiro, 2001; Lerner and Tirole, 2004; Boldrin and Levine, 2005; Llanes and Trento, 2009). Looking at what changes when we add the dynamic dimension is interesting.

To see what happens in the static case, assume only one innovation is under consideration. The innovation uses $n-1$ old ideas, which have already been invented. If the innovation is performed, the innovator obtains a value v from a uniform distribution between 0 and 1, and incurs in a cost ε of R&D. The probability of innovation is $Pr = 1 - \varepsilon - c_n$, with patents or patent pool and $Pr = 1 - \varepsilon/\phi$ without patents.

With patents, the patent holder's problem is to maximize $Pr\, p^i$. As a result, the equilibrium price is $\frac{1-\varepsilon}{n}$ and the probability of innovation is $\frac{1-\varepsilon}{n}$. We have shown that in the dynamic model, the probability of innovation is $\frac{1-\varepsilon}{n} + \beta J_{n+1} - \frac{n-1}{n}\phi\beta J_n$, with $J_n, J_{n+1} > 0$. These extra terms arise because the innovator gets licensing revenues from future innovators. Dynamic incentives imply a higher probability of innovation, but the increase is not enough to prevent the probability of innovation from converging to 0 as $n \to \infty$.

A patent pool would consider cross-price effects, which would lead to a price of $\frac{1-\varepsilon}{2\,(n-1)}$ and a probability of innovation of $\frac{1-\varepsilon}{2}$. The probability of the corresponding dynamic model is $\frac{1-\varepsilon}{2} + \frac{n}{2}\beta J_{n+1} - \frac{n-1}{2}\phi\beta J_n$, with $J_n, J_{n+1} > 0$. In this case, the extra terms arise due to not only the future licensing revenues of the innovator, but also to the pool's concern with keeping the future licensing revenues of current patent holders.

With respect to the no-patents case, the profit-maximizing decision is the same as in the dynamic case. Innovators will therefore perform the innovation if $\phi\, v_n \geq \varepsilon$, which leads to a probability of $Pr = 1 - \varepsilon/\phi$. However, in the dynamic case, innovation is suboptimal even when $\phi = 1$, which contrasts with the static case, where innovation is socially optimal because no intertemporal link between innovations exists and therefore neither does any externality.

10. Finite patents

We have seen that with patents of infinite length, innovation is stifled as n increases due to the complementary monopoly problem. In this section, we ask whether appropriately reducing the length of patents can prevent this problem.[3]

For simplicity, we assume $\phi = 0$ and $\beta = 1$, which is the most favorable case for patents: ideas are scarce and the discount factor is small, so, in principle, society could largely benefit if innovators with a low value for their inventions could get additional revenues from charging other innovators.

Each stage corresponds to one period and only one innovation is attempted at each period. If the innovator decides to introduce the innovation, she obtains a patent for L periods. The innovator, therefore, has to pay licensing fees for L previous innovations, but she also charges licenses to L future innovators.

The main difficulty of the present analysis is that, unlike in the previous sections, the identity of the patent holders matters. The price and future expected licensing revenues will be different for different patent holders, depending on how long her patent lasts.

The innovator will introduce the innovation if the revenues from innovation are larger than the cost:

$$v_n + \sum_{m=n+1}^{n+L} p_m^n \prod_{k=n+1}^{m} Pr_k \geq \sum_{i=n-L}^{n-1} p_n^i + \varepsilon,$$

which means the probability of innovation is

$$Pr_n = 1 + \sum_{m=n+1}^{n+L} p_m^n \prod_{k=n+1}^{m} Pr_k - \sum_{i=n-L}^{n-1} p_n^i - \varepsilon.$$

The L current patent holders differ in their objective functions. Let J_n^i be the future expected revenues of patent holder i at stage n, given that stage n has been reached. Then

$$J_n^i = Pr_n(p_n^i + J_{n+1}^i).$$

The patent holder charging a license for the last time is patent holder $n - L$, so $J_{n+1}^{n-L} = 0$. The patent of $n - L + 1$, on the other hand, will

[3] We have also analyzed the effects of reducing the breadth of patents. For example, suppose new inventions may infringe on old patents with certain probability. Within our framework, the effects of reducing breadth are similar to the effects of reducing patent length: a lower breadth implies that the innovator will have to pay fewer licensing fees, but it also means fewer future inventions will infringe on her patent.

last for one more period, so $J_{n+1}^{n-L+1} = Pr_{n+1}\, p_{n+1}^{n-L+1}$. In this way, we can construct the future expected revenues of the L patent holders.

The profit maximization problem is

$$\max_{p_n^i} J_n^i = Pr_n(p_n^i + J_{n+1}^i).$$

The first-order condition is $-p_n^i - J_{n+1}^i + Pr_n = 0$, so $p_n^i + J_{n+1}^i = Pr_n$ and $J_n^i = Pr_n^2$ for all i, which also implies $p_n^{n-L} = Pr_n$.

We are interested in stationary equilibria, which means $Pr_n = Pr$ for all n. Stationarity, together with the first order condition, implies $p_n^i = Pr(1 - Pr)$ for $i \geq n - L$. Substituting the equilibrium prices into the probability of innovation, we get

$$\begin{aligned} Pr &= 1 + \sum_{m=n+1}^{n+L} p_m^n \prod_{k=n+1}^{m} Pr_k - \sum_{i=n-L}^{n-1} p_n^i - \varepsilon \\ &= 1 + \sum_{m=1}^{L-1} Pr(1-Pr)Pr^m + PrPr^L - (L-1)Pr(1-Pr) - Pr - \varepsilon. \end{aligned}$$

Solving for Pr, we get:

$$Pr = \frac{L + 1 - \sqrt{(L-1)^2 + 4L\varepsilon}}{2L},$$

which is the stationary equilibrium probability of innovation.

Figure 3 shows the probability of innovation as a function of the patent length for $\varepsilon = 0.2$. We can see the probability of innovation decreases with L, which means patents hurt more than benefit the innovator, because the innovator has to pay licenses to the patent holders. Future licensing revenues are uncertain, however, as they depend on future innovations being performed.

Also note that $Pr \to 0$ when $L \to \infty$ and $Pr \to 1 - \varepsilon$ when $L \to 0$, which corresponds to the previously analyzed patents and no-patents cases (with $\theta = 1$).

10.1. **Revenues depend on patent length.** We have assumed that the revenues from selling the new product in the final goods market are independent of patent length. In this subsection, we analyze what happens when we relax this assumption. Assume the revenues of the innovator are $\psi(L)\, v_n$, with $\psi'(L) \geq 0, \psi''(L) \leq 0, \lim_{L\to 0} \psi(L) = \underline{\psi}$ and $\lim_{L\to\infty} \psi(L) = 1$. Here, $\underline{\psi}$ is the fraction of social surplus the innovator would appropriate without any patent protection due to trade secrets or first-mover advantages.

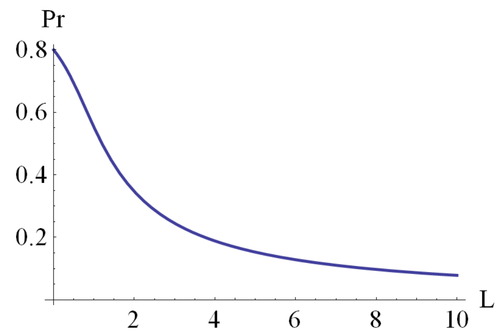


FIGURE 3. Probability of innovation and patent length.

In this case, the innovator will innovate if

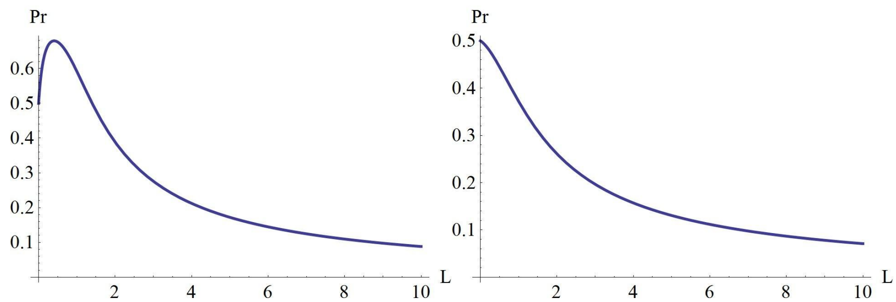
$$\psi(L)\,v_n + \sum_{m=n+1}^{n+L} p_m^n \prod_{k=n+1}^{m} Pr_k \geq \varepsilon + \sum_{i=n-L}^{n-1} p_n^i.$$

Applying a procedure similar to that in the previous case, we obtain the probability of innovation in the stationary equilibrium:

$$Pr = \frac{L+1-\sqrt{(L-1)^2+4L\varepsilon/\psi(L)}}{2L}.$$

The effect of patent length on the probability of innovation depends on the functional form of $\psi(L)$. Let $\psi(L) = 1 - \frac{1-\underline{\psi}}{(L+1)^{\gamma}}$, where γ measures the speed at which revenues grow when L increases. Figure 4a shows that when ψ is more concave ($\gamma = 1$), the probability of innovation first increases and then decreases with patent length. The optimal length is positive and finite (in this case $L = 1$). Figure 4b shows that, for a lower degree of concavity of $\psi(L)$, completely removing patents is optimal. Therefore, the results do not change significantly when the revenues in the final-goods sector depend on patent length.

In this model, short patents therefore perform better than long patents. O'Donoghue et al (1998) find that when patent breadth is infinite, which is always the case in this model, long patents stimulate innovation activity. These apparently different results arise because we are looking for solutions to different problems. We analyze the effect of patent policy on the complementary monopoly problem, disregarding its effect on the final-goods market. O'Donoghue et al (1998) do exactly the opposite. Therefore, our findings are not opposed to theirs, but rather are complementary.



(A) $\underline{\psi} = 0.2, \varepsilon = 0.1, \gamma = 1.$ (B) $\underline{\psi} = 0.2, \varepsilon = 0.1, \gamma = 0.1.$

FIGURE 4. Probability of innovation as a function of patent length.

11. Conclusion

In this paper, we build a dynamic model where the accumulation of patents generates an increasing number of claims on cumulative innovation. The model is intended to reproduce the central feature of innovation activity in high-tech industries: new products are more complex than old products, because they build on a larger stock of previously accumulated knowledge.

We study the policy that maximizes expected social welfare and compare it with the outcome of three patent-policy regimes: patents, patent pools, and no patents. We find that, even abstracting from the monopolistic inefficiencies of patents, none of these policies attains the optimum.

With patents, the innovator has to pay an increasing number of license fees to previous innovator. Asymmetric information on the value of the innovation and uncoordinated market power of licensors create an anticommons effect that reduces the incentives to innovate as innovation becomes more complex. The anticommons effect is weaker than in the static case, but it is still strong enough to drive the probability of innovation to zero as the number of licenses grows. Enforcing a patent pool solves the lack of coordination but not the asymmetric-information problem. As a result, the outcome of patent pools is more desirable but still does not achieve the first best. Eliminating patent protection solves the two problems but introduce a non-internalized externality: previous innovations set the foundations for future innovations. Therefore the social cost of one innovation may be higher than its instantaneous social value (the social value the innovation creates *per se*), and yet the innovation may be socially desirable because it allows the development of further innovations. This is the standard problem of basic research.

Then we study alternative solutions to the anticommons: (i) the complete sale of patent rights of each innovator to the next one and (ii) the possibility that the licensee will compete with the original licensor. Both alternatives exactly replicate the sequence of innovations under the patent-pool regime. Another interesting result of the paper is that patent pools are dynamically unstable, as the incentives to remain outside the pool increase as the sequence of innovations progresses.

We also find the first best can be reached by decentralizing the innovation decision and implementing a tax-subsidy scheme, even if the government and the previous innovators do not know the value of the innovation. Also, when studying optimal patent length, we find that short patents may increase innovation activity by alleviating the complementary monopoly problem.

This paper shows that patent protection may be the wrong way to provide incentives to innovation in complex industries such as electronics, software, and hardware. Enforcing patent pools or eliminating patent protection would improve welfare, but still would not reach the social optimum. We hope this paper will contribute to future research on the design of an optimal innovation policy.

References

Bergstrom TC (1978) Cournot equilibrium in factor markets. Department of Economics, University of California, Santa Barbara

Boldrin M, Levine DK (2005) The economics of ideas and intellectual property. Proceedings of the National Academy of Sciences 102(4):1252–1256

Brenner S (2009) Optimal formation rules for patent pools. Economic Theory 40(3):373–388

Chang HF (1995) Patent scope, antitrust policy, and cumulative innovation. RAND Journal of Economics 26(1):34–57

Chari V, Jones L (2000) A reconsideration of the problem of social cost: Free riders and monopolists. Economic Theory 16(1):1–22

Cournot A (1838) Researches Into the Mathematical Principles of the Theory of Wealth. Irwin (1963)

Erkal N, Scotchmer S (2007) Scarcity of Ideas and Options to Invest in R&D. University of California, Berkeley, Department of Economics, Working Paper 07-348

Garfinkel S, Stallman R, Kapor M (1991) Why patents are bad for software. Issues in Science and Technology 8(1):50–55

Green JR, Scotchmer S (1995) On the division of profit in sequential innovation. RAND Journal of Economics 26(1):20–33

Heller MA (1998) The tragedy of the anticommons: Property in the transition from marx to markets. Harvard Law Review 111(3):621–688

Heller MA, Eisenberg RS (1998) Can patents deter innovation? the anticommons in biomedical research. Science 280(5364):698–701

Holmstrom B (1982) Moral hazard in teams. The Bell Journal of Economics 13(2):324–340

Hopenhayn H, Llobet G, Mitchell M (2006) Rewarding sequential innovators: Prizes, patents, and buyouts. Journal of Political Economy 114(6):1041–1068

Lerner J, Tirole J (2004) Efficient patent pools. American Economic Review 94(3):691–711

Llanes G, Trento S (2009) Anticommons and optimal patent policy in a model of sequential innovation. Working Paper 09-148, Harvard Business School

Menezes F, Pitchford R (2004) A model of seller holdout. Economic Theory 24(2):231–253

O'Donoghue T, Scotchmer S, Thisse J (1998) Patent breadth, patent life, and the pace of technological progress. Journal of Economics & Management Strategy 7(1):1–32

Scotchmer S (1991) Standing on the shoulders of giants: Cumulative research and the patent law. Journal of Economic Perspectives 5(1):29–41

Scotchmer S (1996) Protecting early innovators: Should second-generation products be patentable? RAND Journal of Economics 27(2):322–331

Shapiro C (2001) Navigating the patent thicket: Cross licenses, patent pools, and standard setting. In: Jaffe AB, Lerner J, Stern S (eds) Innovation policy and the economy, vol 1, MIT Press for the NBER, pp 119–150

Sonnenschein H (1968) The dual of duopoly is complementary monopoly: or, two of cournot's theories are one. The Journal of Political Economy 76(2):316–318