# International Protection of Intellectual Property*

by

Gene M. Grossman
Princeton University

and

Edwin L.-C. Lai
City University of Hong Kong

September 2002

**Abstract**

We study the incentives that governments have to protect intellectual property in a trading world economy. We consider a world economy with ongoing innovation in two countries that differ in market size and in their capacities for innovation. We associate the strength of IPR protection with the duration of a country's patents that are applied with national treatment. After describing the determination of national policies in a non-cooperative regime of patent protection, we ask, Why are patents longer in the North? We also study international patent agreements by deriving the properties of an efficient global regime of patent protection and asking whether harmonization of patent policies is necessary or sufficient for global efficiency.

**JEL Classification:** O34, F13

**Keywords:** patents, intellectual property, harmonization, TRIPs

*The authors thank Elhanan Helpman, Robert Lucas, Jr., Giovanni Maggi, Keith Maskus and Lars Svensson for useful discussions and comments. They are grateful to the U.S. National Science Foundation, the Hong Kong Research Grants Council (Project no. CityU 1145/99H), and the Department of Economics and Finance at City University of Hong Kong for financial support.

# 1 Introduction

During the 1980's and early 1990's, the United States and several European countries expressed strong dissatisfaction with what they deemed to be inadequate protection of intellectual property in many developing countries. The developed countries made the upgrading of intellectual property rights (IPR) protection one of their highest priorities for the Uruguay Round of trade talks. Their efforts in those negotiations bore fruit in the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs), which was approved as part of the Final Act of the Uruguay Round.

The TRIPs agreement establishes minimum standards of protection for several categories of intellectual property. For example, in the area of new technology, it requires countries to grant patents to a broad class of innovations for a minimum of twenty years and to treat foreign and domestic patent applicants alike. But IPR protection remains a highly contentious issue in international relations between the North and the South, because many developing countries believe that the TRIPs agreement was forced upon them by their economically more powerful trading partners and that this move toward harmonization of patent policies serves the interests of the North at the expense of those of the South.

In a country that is closed to international trade, the design of a system of IPR protection poses a clear trade-off to a welfare-maximizing government. By strengthening the protection of intellectual property, a government provides greater incentives for innovation and thus the benefits that come from having more and better products. But, at the same time, it curtails potential competition for firms that have previously innovated and thus limits the benefits that can be realized from existing products. As Nordhaus (1969) argued, the optimal patent policy equates the marginal dynamic benefit with the marginal static efficiency loss.

But in an open economy, the trade-offs are not so clear cut. International trade spreads the benefits of innovation beyond national boundaries. This means that a country does not reap all of the global benefits that come from protecting intellectual property within its borders. Moreover, countries differ in their capacities for innovation due to differences in skill endowments and technical know-how. It is not obvious how a government ought to set its national IPR policy if some of the benefits of its national innovation accrue to foreigners, if its constituents benefit from innovations that are encouraged and take place beyond its boundaries, and if domestic and foreign firms differ in their ability to innovate.

Some previous research has addressed the question of whether a country with a limited capacity to innovate will benefit from extending IPR protection to foreign inventors. Chin and Grossman (1990) and Deardorff (1992) investigated the welfare effects of extending patent protection from the country in which innovation takes place to another country that only consumes the innovative products. Both of these papers treat the investment in R&D as a once-off decision. In contrast, Helpman (1993) models innovation as an ongoing process and associates the strength of the IPR regime with the flow probability that a given product protected by a patent in the North will be imitated in the South. He evaluates the welfare consequences of marginal changes in the rate of imitation. These papers do not, however, consider the simultaneous choice of IPR protection by trade partners, nor do they discuss what international regime of IPR protection would be globally efficient.[1]

In this paper, we study the incentives that governments have to protect intellectual property in a trading world economy. We consider a world economy with ongoing innovation in which there are two countries that differ in market sizes and in their capacities for conducting research and development. Innovators develop the designs for new products, each of which has a limited economic life. We associate the strength of IPR protection with the duration of a country's patents. Patents provide inventors with exclusive rights to produce, sell and distribute their products within a country. We study a regime with national treatment, which means that the same protection is provided to all inventors regardless of their national origin.

We begin in Section 2 with the case of a closed economy. There we re-examine the trade-off between static costs and dynamic benefits that was first studied by Nordhaus. We derive a neat formula that characterizes the optimal patent policy in

[1] McCalman (1997) addresses some of these issues in a model of once-off innovation by a single firm in a developed economy.

a closed economy, and discuss the determinants of the optimal patent length. One interesting finding is that the optimal duration of patents may be independent of or even decreasing in the size of the economy.

In Section 3, we describe the determination of national policies in a non-cooperative regime of patent protection. We derive best response functions for the "North" and the "South," where the North is assumed to have a higher wage than the South, as well as possibly a larger market for innovative products and a greater capacity for innovation. The best response is a patent length that maximizes a country's national welfare, given the duration of patents in its trading partner. We characterize the best responses, compare the incentives for providing protection for intellectual property in an open economy to those that exist in a closed economy, and explain the strategic interactions between countries in the setting of their patent policies.

In Section 4, we ask, Why are patents longer in the North? If the capacity for R&D greater in the North than in the South and the market for innovative products is at least as large there, then patent duration will be longer in the North than in the South in a Nash equilibrium. We explain why relative market size and relative productivity in innovation matter for the relative incentives to protect intellectual property. Patents are a more potent instrument for stimulating innovation in the relatively larger market. And a country that invents a smaller share of the world's innovative products will find more incentive to curtail patent protection so as to benefit local consumers at the expense of producers.

We study international patent agreements in Section 5. First we derive the properties of an efficient global regime of patent protection. An efficient patent regime is one that provides the optimal aggregate incentives for innovation to inventors throughout the world. These incentives can be achieved by various combinations of patent policies in the two countries, so there is no unique pair of patent lengths that is needed for global efficiency. However, different ways of achieving the optimal aggregate incentives have different implications for the distribution welfare between the North and the South. Among combinations of policies that give the same overall incentives for global research, the North fares better, and the South worse, the longer are patents in the South. An implication of our findings is that harmonization of patent policies is neither necessary nor sufficient for global efficiency. Moreover, starting from a non-cooperative equilibrium with longer patents in the North than in the South, an efficient agreement calling for harmonization of patent lengths benefits the North quite possibly at the expense of the South.

Readers familiar with the literature on trade policy will recognize a familiar structure in our inquiry. Our examination of a non-cooperative regime of patent protection is analogous to Johnson's (1953-54) study of non-cooperative tariff setting by two large countries. Our subsequent identification of the efficient combinations of patent policies is analogous to Mayer's (1981) similar examination of the efficient combinations of trade policies. We, like Mayer, associate the efficiency frontier with the possible outcomes of an international negotiation.

In Section 6, we extend our analysis of both the non-cooperative and cooperative settings to a world with many trading countries. The many country model is qualitatively similar to the two-country model, although the addition of more countries exacerbates the inefficiencies associated with non-cooperation. Our findings are summarized in Section 7.

## 2 A Simple Model of Innovation

In this section, we construct a simple model of ongoing innovation. We develop the model for a closed economy and use it to revisit the question of the optimal patent length that was first addressed by Nordhaus (1969). Our model yields a neat formula that characterizes the trade-off between the static costs and dynamic benefits of extending the period of patent protection. The discussion of a closed economy lays the groundwork for the more subtle analysis of the international system that we undertake in the sections that follow.

The economy has two sectors, one that produces a homogeneous good and another that produces a continuum of differentiated products. The designs for the differentiated products result from private investments in R&D. Once a good has been invented, it has a finite economic life of length $\bar{\tau}$. That is, a new product potentially provides utility to consumers for a period of $\bar{\tau}$ from the time of its creation, whereupon its value to consumers drops to zero.

There are $M$ consumers with identical preferences. We shall refer to $M$ as the "size of the market."[2] The representative consumer maximizes a utility function of the form

$$U(t) = \int_t^\infty u(z)e^{-\rho z}dz \tag{1}$$

where

$$u(z) = y(z) + \int_0^{n(z)} h[x(i,z)]di, \tag{2}$$

$y(z)$ is consumption of the homogeneous good at time $z$, $x(i,z)$ is consumption of the $i^{th}$ variety of differentiated product at time $z$, and $n(z)$ is the measure of differentiated products invented before $z$ that still hold value to consumers at time $z$. We assume that $h'(x) > 0$, $h''(x) < 0$, $h'(0) = \infty$, and $-xh''(x)/h'(x) < 1$ for all $x$. The third assumption ensures a positive demand for every variety at any finite price. The fourth ensures that any firm holding a patent for a differentiated product will charge a finite price.

A consumer maximizes utility by purchasing some of all varieties that are not yet obsolete. He chooses $x(i,z)$ so that $h'[x(i,z)] = p(i,z)$ for all $i$ and $z$, where $p(i,z)$ is the price of variety $i$ at time $z$. After the consumer makes all of his optimal purchases of differentiated products at time $z$, he devotes the remainder of his spending to the homogeneous good $y$. Spending is always positive in the equilibria we describe. This means that the interest rate is constant and equal to $\rho$, from the condition for intertemporal optimization.

Manufacturing requires only labor. Any firm can produce good $y$ with $a$ units of labor per unit of output. All known varieties of the differentiated product also can

[2] In our model, demand for differentiated products does not vary with income. Thus, a rich country need not have a larger market for these goods than a poor country. Nonetheless, we prefer to think of the market for differentiated goods as being larger in the North than in the South. This could be rigorously justified within our model if we were to suppose that differentiated products provide utility only after a threshold level of consumption of the homogeneous goods has been reached. Then, a rich country is likely to have more consumers who surpass the threshold.

be produced with $a$ units of labor per unit of output. But the government grants the original designer of a differentiated product the sole rights of production and sale for a period of length $\tau$. We assume that patents are perfectly enforced.

The design of new varieties requires both labor and human capital. For simplicity, we take $\phi(z) \equiv F[H, L_R(z)] = \{b\,[L_R(z)/a]^{\beta} + (1-b)H^{\beta}\}^{1/\beta}$, where $\phi(z)$ is the flow of new inventions at time $z$, $H$ is the (constant) stock of human capital, $L_R(z)$ is the amount of labor devoted to R&D, and $a$ is a measure of labor productivity as before. This is, of course, a production function with a constant elasticity of substitution between labor and human capital. We assume that $\beta \leq 1/2$, or equivalently that the elasticity of substitution is less than or equal to two. This assumption is sufficient (but not necessary) to ensure that any patent length that satisfies the first-order condition for an interior optimum also satisfies the second-order condition. Note that $\dot{n}(z) = \phi(z) - \phi(z - \bar{\tau})$, because the goods that were invented at time $z - \bar{\tau}$ become obsolete at time $z$.

We describe now the static and dynamic equilibrium for an economy that has a patent duration of $\tau$. In equilibrium, firms with live patents for differentiated products behave as monopolies. Each such firm faces an inverse demand curve from each of the $M$ consumers with the form $p(x) = h'(x)$. The firm sets its price so that $(p - aw)/p = -xh''/h'$, where $w$ is the wage rate and $x$ is sales per consumer. This is the usual monopoly-pricing rule whereby the markup over unit cost as a fraction of the price is equal to the inverse demand elasticity. Optimal pricing yields a typical patent holder profits of $\pi$ per consumer, and total profits of $M\pi$.

When a patent expires, competitors can imitate the good costlessly. Then the product sells for the competitive price of $p = aw$ and generates no further profits. This pricing of the good continues until the good becomes obsolete. Meanwhile, the homogeneous good always carries the competitive price of $aw$, which, because this good is the numeraire, implies that $w = 1/a$. In writing this condition, we implicitly assume that the economy's labor supply is sufficiently large that some labor remains for production of the homogeneous good after all derived demand for labor for producing differentiated products and conducting R&D has been satisfied.

Labor engages in manufacturing and R&D. The labor employed in manufacturing differentiated goods is just the amount needed to produce the quantities demanded at the equilibrium prices. The allocation of labor to R&D is such that its marginal value product in this activity is equal to the wage rate. Thus,

$$vF_L\left(H, L_R\right) = w, \tag{3}$$

where $v$ is the value of a new patent. Since there is no uncertainty about future earnings, patents are worth the discounted value of the profits they generate in the time before they expire, or

$$v = \frac{M\pi}{\rho}\left(1 - e^{-\rho\tau}\right). \tag{4}$$

We can see from (3) and (4) that an increase in the patent length increases the value of a new patent, thereby drawing additional resources into R&D.

The final equilibrium condition equates savings with investment. Savings are the difference between national income $rH + wL + n_m M\pi$ and aggregate spending $E$, where $r$ is the return to human capital, $L$ is the aggregate labor supply, and $n_m$ is the number of differentiated products that retain their patent protection. All investment is devoted to R&D. This activity has an aggregate cost of $rH + wL_R$. Thus, we can write the equilibrium condition as $(rH + wL + n_m M\pi) - E = rH + wL_R$, or

$$E = w(L - L_R) + n_m M\pi. \tag{5}$$

It is useful to calculate an expression for aggregate welfare at date 0, the time at which a new (optimal) patent policy will be set by the government. By assumption, this patent protection applies only to goods introduced after time 0; those introduced beforehand are subject to whatever policy was in effect at the time of their invention.[3] At each moment in time, each of the $M$ consumers enjoys surplus of $C_m = h(x_m) - p_m x_m$ from his consumption of any good under patent. Here, $x_m$ is the amount sold

[3]It would never be optimal for the government to extend patent protection on goods that have already been invented. This would create deadweight loss without any offsetting social benefit. The government might wish to eliminate protection for goods that were invented under a different regime, but we assume that such expropriation of intellectual property would not be legal.

by the typical monopoly to the typical consumer and $p_m$ is the monopoly price. We distinguish between those goods invented before time 0 and those invented afterward. The former yield some exogenous surplus that is unaffected by the new patent policy. Of the latter, there are $s\phi$ at time $s$, for $s$ between 0 and $\tau$, and a constant number $\tau\phi$ thereafter. Each consumer also enjoys surplus of $C_c = h(x_c) - p_c x_c$ from his purchases of any competitively-priced variety of the differentiated product, where $x_c$ and $p_c$ are the quantity and price of a typical one of these purchases. Again, the competitively-priced goods that were invented before time 0 yield some exogenous surplus. The number of such goods invented after time 0 that are still economically viable at time $s$ is 0 for $s \leq \tau$, $(s-\tau)\,\phi$ for $s$ between $\tau$ and $\bar{\tau}$, and $(\bar{\tau}-\tau)\,\phi$, for $s \geq \bar{\tau}$. Using (1), (2) and (5), we calculate that utility at time 0 is

$$U(0) = \Lambda_0 + \frac{w\,(L-L_R)}{\rho} + \frac{M\phi}{\rho}(C_m+\pi)T + \frac{M\phi}{\rho}C_c(\bar{T}-T) \qquad (6)$$

where $\Lambda_0$ is the discounted present value of the consumer surplus and profits derived from goods invented before time 0, and where $T \equiv \left(1-e^{-\rho\tau}\right)/\rho$ and $\bar{T} \equiv \left(1-e^{-\rho\bar{\tau}}\right)/\rho$. Note that $T$ is the present discounted value of a flow of one dollar from time 0 to time $\tau$, and that $\bar{T}$ has an analogous interpretation.

We are now ready to derive the optimal patent length for a closed economy. Formally, we maximize $U(0)$ with respect to $\tau$, after recalling that $\phi = F(H, L_R)$ and that $L_R$ is a function of $\tau$ via (3) and (4).[4] It is more intuitive, however, to describe the social costs and benefits that derive from extending the patent length marginally from a given length $\tau$. The cost of lengthening the period of patent protection is that the economy suffers the deadweight loss of $M\,(C_c - C_m - \pi)$ on each of the differentiated products invented after time 0 for a marginally longer period of time. If the patent period is lengthened at time 0, the extra deadweight loss kicks in at time $\tau$, and continues thereafter. The flow of new products is $\phi$ per unit time. Thus,

[4] Equivalentlly, we can maximize $\rho U(0)$ over the choice of $T$. Note that $C_m$, $C_c$ and $\pi$ do not depend on the duration of patents and thus do not depend on $T$. We can combine (3) and (4) to write $M\pi T F_L\,(H, L_R) = w$, which allows us to solve for the functional relationship betwen the labor devoted to R&D and the policy variable $T$; denote it by $L_R(T)$. Then, substituting this expression into (6) and rearranging terms, we can write the maximand as

the total marginal cost, discounted to time 0, is

$$\frac{\phi e^{-\rho\tau}}{\rho} M\left(C_c - C_m - \pi\right).$$

The benefit to the economy of extending the patent length is that it encourages R&D, which in turn means a greater variety of differentiated products. Each differentiated product yields discounted consumer surplus of $MC_mT$ over its life as a patented product and $MC_c(\bar{T} - T)$ over its life as a competitively-priced product, where in each case the discounting is back to the time of invention. Now if we discount this flow of benefits back to time 0, and multiply by the number of new inventions induced by a marginal lengthening of the patent period, we have the total marginal benefit, which is equal to

$$\frac{1}{\rho} \cdot \frac{d\phi}{dv} \cdot \frac{dv}{d\tau} \cdot \left[MC_mT + MC_c(\bar{T} - T)\right].$$

Using (3) we calculate that

$$\frac{d\phi}{dv} = \gamma\frac{\phi}{v},$$

where $\gamma$ is the ratio of the elasticity of research output with respect to labor to the elasticity of the marginal product of labor in R&D; i.e., $\gamma \equiv -\left(F_L\right)^2 / \left(FF_{LL}\right)$. The variable $\gamma$ identifies the responsiveness of innovation to the protections afforded by the patent system. In general, it is a function of $L_R$ and thus indirectly of the patent length $\tau$. With the CES research technology, $\gamma = [b/(1-b)(1-\beta)](L_R/aH)^{\beta}$. For the special case of a Cobb-Douglas technology (which is the limiting case of the CES as $\beta \to 0$), $\gamma$ is a constant equal to the ratio of the cost share of labor to the cost share of human capital.

---

$$\rho U(0) = \rho\Lambda_0 + w\left[L - L_R(T)\right] + MF\left[H, L_R(T)\right]\left[\left(C_m + \pi - C_c\right)T + C_c\bar{T}\right].$$

The first-order condition for a maximum requires

$$(C_c - C_c - \pi)MF\left[H, L_R(T)\right] = \left\{MF_L\left[\left(C_m + \pi - C_c\right)T + C_c\bar{T}\right] - w\right\}L'_R$$

from which (7) follows. In the appendix we show $\beta \leq 1/2$ is sufficient to enure that the second order condition is satisfied at any value of $T$ that satisfies the first order condition (7).

Next, we use (4) to compute that

$$\frac{dv}{d\tau} = M\pi e^{-\rho\tau}.$$

Substituting for $d\phi/dv$ and $dv/dt$ in the expression for marginal benefit, and equating the result to the marginal cost, we derive an implicit formula for the optimal patent length. We find that

$$C_c - C_m - \pi = \gamma \left[ C_m + C_c \left( \frac{\bar{T} - T}{T} \right) \right] \tag{7}$$

at an (interior) optimal value of $T$.

From (7) we see that the optimal patent is longer, the greater is the useful life of a product (larger $\bar{\tau}$), the more patient are consumers (smaller $\rho$), and the greater is the ratio of consumer surplus plus profits under monopoly to consumer surplus with competition.[5] All of these findings accord well with intuition. One noteworthy feature of (7) is the relationship between market size and the optimal patent length. In a closed economy, the first-best level of R&D — that which maximizes discounted utility when all goods are competitively priced — typically is an increasing function of market size. This is because innovation is a public good, and the Samuelsonian rule for optimal provision of a public good calls for greater output when the benefits can be spread across more consumers. But the encouragement of innovation by patents achieves only a second best. According to (7), the size of the market $M$ affects the optimal patent length only through its effect on the supply elasticity of innovations. If $\gamma$ is an increasing function of $L_R$, as it will be if $1/2 > \beta > 0$, then the optimal $\tau$ is an increasing function of $M$. But if $\gamma$ is a decreasing function of $L_R$, as it will be if $\beta < 0$, then $\tau$ is an decreasing function of $M$. In the benchmark Cobb-Douglas case (with $\beta = 0$), $\gamma$ is independent of $L_R$ and therefore of market size. Then an increase in $M$ enhances both the marginal benefit of extending patents and the marginal cost of doing so, but does so in equal proportions. The optimal patent length in a closed economy with a Cobb-Douglas research technology is invariant to market size.

[5] The proof of these statements makes use of the second-order condition, which ensures that the right-hand side of (7) is a declining function of $T$.

# 3 Noncooperative Patent Protection

In this section, we study the national incentives for protection of intellectual property in a world economy with imitation and trade. We derive the Nash equilibria of a game in which two countries set their patent policies simultaneously and noncooperatively. The countries are distinguished by their wage rates, their market sizes, and their stocks of human capital. The last of these proxies for their different capacities for R&D. We shall term the countries "North" and "South," in keeping with our desire to understand the tensions that surrounded the tightening of IPR protection in the developing countries in the last decade. Maskus (2000a, ch.3) has documented an increase in innovative activity in poor and middle-income countries such as Brazil, Korea, and China, so our model of relations between trading partners with positive but different abilities to conduct R&D may be apt for studying the incentives for IPR protection in a world of trade between such nations and the developed economies.[6] But our model may apply more broadly to relations between any groups of countries that have different wages and different capacities for research. Such differences exist, albeit to a lesser extent than between North and South, in the comparison of countries in Northern and Southern Europe, or the comparison of the United States and Canada. We do not mean the labels North and South to rule out the application of our analysis to these other sorts of relationships.

## 3.1 The Global IPR Regime

The model is a natural extension of the one presented in Section 2. Consumers in the two countries share identical preferences. In each country, the representative consumer maximizes the intertemporal utility function in (1). The instantaneous utility of a consumer in country $j$ now is given by

$$u_j(z) = y_j(z) + \int_0^{n_S(z)+n_N(z)} h[x_j(i,z)]di, \tag{8}$$

[6] He also shows the extent to which patent applications in countries like Mexico, Brazil, Korea, Malaysia, Indonesia and Singapore are dominated by foreign firms, a feature of the data that figures in our analysis.

where $y_j(z)$ is consumption of the homogeneous good by a typical resident of country $j$ at time $z$, $x_j(i,z)$ is consumption of the $i^{th}$ differentiated product by a resident of country $j$ at time $z$, and $n_j(z)$ is the number of differentiated varieties previously invented in country $j$ that remain economically viable at time $z$. There are $M_N$ consumers in the North and $M_S$ consumers in the South. While we do not place any restrictions on the relative sizes of the two markets at this juncture, we shall be most interested in the case where $M_N > M_S$.[7] It does not matter for our analysis whether consumers can borrow and lend internationally or not.

In country $j$, it takes $a_j$ units of labor to produce one unit of the homogeneous good or to produce one unit of any variety of the differentiated product. Of course, the rights to produce some varieties may be limited by ongoing patent protection in one or both countries. New goods are invented in each region according to $\phi_j = F(H_j, L_{Rj}/a_j) = \left[b[L_{Rj}/a_j]^\beta + (1-b)H_j^\beta\right]^{1/\beta}$, where $H_j$ is the human capital endowment of country $j$, $L_{Rj}$ is the labor devoted to R&D there, and again we take $\beta \leq 1/2$. We assume that $a_N < a_S$, which means that labor is uniformly more productive in the North than in the South.[8] This implies $w_N/w_S = a_S/a_N > 1$; i.e., the wage in the North exceeds that in the South. We also assume that the numeraire good is produced in positive quantity somewhere in the world economy, so that $w_j = 1/a_j$ for $j = S, N$.

We now describe the IPR regime. In each country, there is *national treatment* in the granting of patent rights. Under national treatment, the government of country $j$ grants a patent of length $\tau_j$ to all inventors of differentiated products regardless of their national origins. In other words, we assume that foreign firms and domestic firms have equal standing in applying for patents in any country. National treatment is required by the TRIPs agreement and it characterized the laws that were in place

[7] We remind the reader that market size is meant to capture not the population of a country, but rather the scale of its demand for innovative products.

[8] In Grossman and Lai (2002), we allowed the relative productivity of labor to vary in different uses; i.e., we allowed for Ricardian comparative advantage across the regions. This feature caused some subtle complications that we ignore here for the sake of simplicity and greater clarity.

in most countries even before this agreement.[9] In our model, a patent is an exclusive right to make, sell, use, or import a product for a fixed period of time (see Maskus, 2000a, p.36). This means that, when good $i$ is under patent protection in country $j$, no firm other than the patent holder or one designated by it may produce the good in country $j$ for domestic sale or for export, nor may the good be imported into country $j$ from an unauthorized producer outside the country. We also rule out parallel imports — unauthorized imports of good $i$ that were produced by the patent holder or its designee, but that were sold to a third party outside country $j$.[10] When parallel imports are prevented, patent holders can practice price discrimination across national markets.

We solve the Nash game in which the governments set their patent policies once-and-for-all at time 0. These patents apply only to goods invented after time 0; goods invented beforehand continue to receive the protections afforded at their times of invention. So long as the governments cannot curtail patents that were previously awarded, the economy has no state variables that bear on the choice of optimal patent policies at a given moment in time. This means that the Nash equilibrium in once-and-for-all patents is also a sub-game perfect equilibrium in the infinitely repeated game in which the governments can change their patent policies periodically, or even

[9] National treatment is required by the Paris Convention for the Protection of Industrial Property, to which 127 countries subscribed by the end of 1994 and 162 countries subscribe today (see http://www.wipo.org/treaties/ip/paris/paris.html). There were, however, allegations from firms in the United States and elsewhere that prior to the signing of the TRIPS agreement in 1994, nondiscriminatory laws did not always mean nondiscriminatory practice. See Scotchmer (2001) for an analysis of the incentives that countries have to apply national treatment in the absence of an enforcible agreement.

[10] The treatment of parallel imports under the TRIPs agreement remains a matter of legal controversy. Countries continue to differ in their rules for territorial exhaustion of IPRs. Some countries, like Australia and Japan, practice international exhaustion, whereby the restrictive rights granted by a patent end with the first sale of the good anywhere in the world. Other countries or regions, like the United States and the European Union, practice national or regional exhaustion, whereby patent rights end only with the first sale within the country or region. Under such rules, patent holders can prevent parallel trade. See Maskus (2000b) for further discussion.

continuously. Of course, the repeated game may have other equilibria in which the governments base their policies at a point in time on the history of policies that were chosen previously. We do not investigate such equilibria with tacit cooperation here, but rather postpone our discussion of cooperation until Section 5.

Let us describe, for given patent lengths $\tau_N$ and $\tau_S$, the life cycle of a typical differentiated product. When a firm invents a new product, it immediately files for patent protection in both countries. This is because there is no cost of the application process and no benefit from waiting to introduce a new good. Moreover, postponing production would generate a loss of value in view of the positive interest rate, while producing without the benefit of patent protection would spell immediate imitation and a loss of profits.

During an initial phase after the product is introduced, the inventor holds an active patent in both countries. Then the patent holder earns a flow of profits $M_N\pi$ from sales to consumers in the North and $M_S\pi$ from sales to consumers in the South, where $\pi$ is earnings per consumer for a monopoly selling a typical brand. Notice that profits per consumer are the same for sales in both markets, because consumers share identical preferences. Also, they do not depend on where a good was invented or where it is produced, because the productivity gap between the countries exactly offsets the wage differential.[11] Households in the North realize a flow of consumer surplus of $M_N C_m$ from their purchases of a typical patented product, while those in the South realize a flow of surplus of $M_S C_m$, where $C_m$ is the surplus enjoyed by a typical consumer of a good produced at a cost of $w_j a_j = 1$ and sold at the monopoly price.

After a while, the patent will expire in one country. For concreteness, let's say that this happens first in the South. Then the good will be imitated by competitive firms producing there, for sales in the Southern market. The imitators will not, however, be able to sell the good in the North, because the live patent there affords protection

[11] In Grossman and Lai (2002), where we allowed for comparative technological advantages across different uses of labor, we were forced to consider separately situations in which direct foreign investment is and is not a possibility. However, our conclusions were qualitatively similar for the two cases.

from such competitive imports. At the moment the patent expires in the South, the price of the good falls there to $w_S a_S = 1$, and the original inventor ceases to realize profits in that market. The flow of consumer surplus in the South rises to $M_S C_c$, where $C_c$ denotes the consumer surplus generated per consumer by a product that is sold for the competitive price of $p_c = 1$. But profits and surplus in the Northern market remain for a while as before.

Eventually, the inventor's patent will expire in the North. Then the Northern market can be served by competitive firms producing in either location. At this time, the price of the good in the North falls to $p_c = 1$ and households there begin to enjoy the higher flow of consumer surplus $M_N C_c$. The original inventor loses his remaining source of monopoly income. Finally, after a period $\bar{\tau}$ has elapsed from the moment of invention, the good becomes obsolete and all flows of consumer surplus cease.

## 3.2 The Best Response Functions

We are now ready to derive the best response functions for the two countries. The best response for a country expresses the patent length that maximizes its aggregate welfare as a function of the given patent policy of its trading partner. Consider the choice of $\tau_S$ by the government of the South. This country bears two costs from prolonging its patents slightly. First, it extends the period during which the country suffers a static deadweight loss of $C_c - C_m - \pi$ per consumer on each good invented in the South. Second, it prolongs the period during which each of its consumers realizes surplus of only $C_m$ instead of $C_c$ on each good that was invented in the North. Notice that the profits earned by Northern producers in the South are not an offset to this latter marginal cost, because they accrue to patent holders in the North. The marginal benefit that comes to the South from prolonging its patents reflects the increased incentive that Northern and Southern firms have to engage in R&D. If the welfare-maximizing $\tau_S$ is positive and less than $\bar{\tau}$, then the marginal benefit per consumer of increasing $\tau_S$ must match the marginal cost, which implies

$$\phi_S(C_c - C_m - \pi) + \phi_N(C_c - C_m) = \frac{\gamma_S \phi_S + \gamma_N \phi_N}{v} M_S \pi \left[ C_m T_S + C_c(\bar{T} - T_S) \right], \quad (9)$$

where $v = (M_S T_S + M_N T_N)\pi$ is the value of a new patent, $T_j = (1 - e^{-\rho\tau_j})/\rho$, and $\gamma_i$ is the responsiveness of innovation in region $i$ to changes in the value of a patent (in elasticity form).

Similarly, in the North, the marginal benefit of extending longer patent protection must match the marginal cost at any interior point on the best response curve. The marginal cost in the North is different from that in the South, because the North's national income includes the profits earned by Northern patent holders but not those earned by Southern patent holders. The marginal benefit differs too, because the effectiveness of patent policy as a tool for promoting innovation varies according to the importance of a country's market in the aggregate profits of potential innovators and because the surplus from a typical product over its lifetime depends upon a country's patent length. The condition for the best response of the North, analogous to (9) above, is

$$\phi_S(C_c - C_m) + \phi_N(C_c - C_m - \pi) = \frac{\gamma_S \phi_S + \gamma_N \phi_N}{v} M_N \pi \left[C_m T_N + C_c(\bar{T} - T_N)\right]. \quad (10)$$

Noting that $\gamma_S = \gamma_N = \gamma$[12], the two best response functions can be written similarly as

$$C_c - C_m - \mu_i \pi = \gamma \frac{M_i T_i}{M_S T_S + M_N T_N} \left[C_m + C_c\left(\frac{\bar{T} - T_i}{T_i}\right)\right] \quad \text{for } i = S, N, \quad (11)$$

[12] The fact that the two supply elasticities $\gamma_S$ and $\gamma_N$ are equal despite the differences in human capital endowments, in employment, and in labor productivity is a property of the CES research technology. It follows from the observation that

$$\gamma_i = \frac{b}{(1-b)(1-\beta)} \left(\frac{L_{Ri}}{a_i H_i}\right)^{\beta}$$

and $vF_L(L_{Ri}/a_i, H_i) = w_i$, or

$$\frac{vb}{a_i}\left[b + (1-b)\left(\frac{L_{Ri}}{a_i H_i}\right)^{-\beta}\right]^{\frac{1-\beta}{\beta}} = \frac{1}{a_i}.$$

Combining the two, we find $\gamma_S = \gamma_N = \gamma$, where

$$\gamma = \frac{b}{1-\beta}\left[\left(\frac{1}{bv}\right)^{\frac{\beta}{1-\beta}} - b\right]^{-1}.$$

where $\mu_i = \phi_i/(\phi_S + \phi_N)$ is the share of world innovation that takes place in country $i$. This form of the best response function facilitates a comparison of the incentives that the government has for protecting intellectual property in a world with trade compared to those that exist when there is no trade, as expressed in equation (7). On the left-hand side of (11), the government of a trading economy considers only a fraction of the profits that flow to patent holders to be an offset to the static cost of continuing patent protection. On the right-hand side, the ability of a trading economy to stimulate innovation with a given change in patent duration is only a fraction of what it is in a closed economy, because inventors earn only part of their discounted profits within the country's borders. Both of these forces point to shorter patent duration in an open economy than would be optimal in the absence of trade. Against this, possibly, is the difference between the supply elasticities for innovation in the closed and open economies; the presence of a foreign country offering patent protection for innovations may increase the responsiveness of innovation to home patent policy if $\gamma$ is an increasing function of $L_R$. However, with the CES research technology $\gamma$ is in fact a *non-increasing* function of $L_R$ whenever $\beta \leq 0$; i.e., when the elasticity of substitution between human capital and labor is less than or equal to one. In such circumstances, the government of an open economy necessarily chooses a shorter duration of patents than it would choose autarky.[13]

In Figure 1, we depict the best response functions in the space of $T_S$ and $T_N$ for the case in which the research technology has a Cobb-Douglas form (i.e., $\beta = 0$). Under such conditions, the supply elasticity $\gamma$ is a constant equal to $b/(1-b)$. Moreover, $\mu_i = H_i/(H_S + H_N)$ for any CES research technology.[14] Thus, both $\mu_i$ and $\gamma$ are

[13]Suppose the government of an open economy were to choose the autarky duration of patents. The marginal cost of extending patents would be greater in the open compared to the closed economy, since $\mu_i < 1$. And, since $\gamma'(L_R) \leq 0$ and $M_i T_i/(M_S T_S + M_N T_N) < 1$, the marginal benefit from extending the patents would be smaller in the open as compared to the closed economy. Thus, the marginal cost would exceed the marginal benefit in the open economy, which means that the government would have reason to cut the patent duration from the autarky level.

[14]Note that

$$\phi_i = H_i \left[ b \left( \frac{L_{Ri}}{a_i H_i} \right)^{\beta} + (1-b) \right]^{\frac{1}{\beta}}.$$

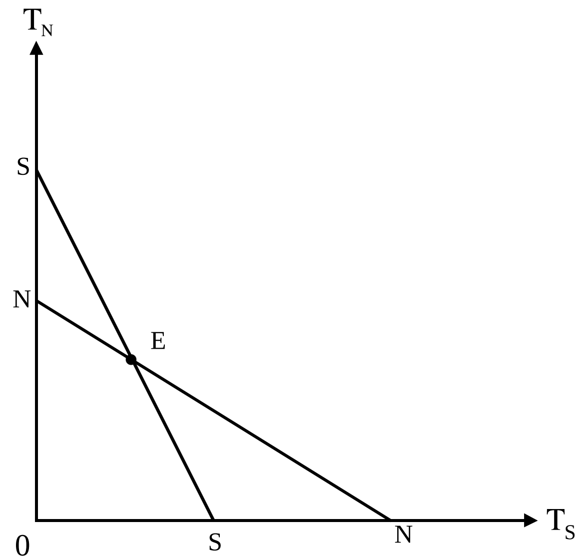


Figure 1: Nash Equilibrium with Cobb-Douglas Research Technologies

independent of the patent policies in the Cobb-Douglas case. It follows from (11) that the best response functions are linear and downward sloping and that the $SS$ curve (the best response function for the South) is steeper than the $NN$ curve (the best response function for the North).

More generally, the best response functions need not be linear in $(T_S, T_N)$ space, but they must be downward sloping whenever $\beta \leq 0$; i.e., when the elasticity of substitution between human capital and labor in designing new products is less than or equal to one. Thus, the patent policies of the two countries are strategic substitutes in such circumstances. To understand the strategic interdependence between the governments in choosing their patent lengths, consider the choice of patent policy by the South. Suppose the North were to lengthen the duration of its patents; i.e., to increase $T_N$. This would shrink the fraction of total discounted profits that an innovator earns in the South and so, ceteris paribus, reduce the responsiveness of

From the fact that $vF_L(L_{Ri}/a_i, H_i) = w_i$, we have that $L_{Ri}/a_iH_i$ takes on a common value in the two countries; see footnote 12. It follows that $\phi_i$ is proportional to $H_i$, with the same factor of proportionality in both countries.

global innovation to the length of Southern patents. Moreover, the increase in $T_N$ would draw labor into R&D in the North and the South. If $\beta < 0$, the elasticity of innovation with respect to patent value would fall. The South would find that its market is relatively less important to potential innovators, and that these innovators are less responsive to patent policy. For both reasons, the marginal benefit to the South of extending patent length would fall at the initial $T_S$, and the government would respond to the increase in $T_N$ with a cut in the duration of its patents.

A situation of strategic complementarity (i.e., upward-sloping best response function) can arise only if the supply elasticity of R&D rises as the size of the research sector expands ($\beta > 0$) and then only if it rises sufficiently much to compensate for the decline in relative importance of a country's market that results when its trading partner lengthens its patents. It is straightforward to show that the two best response functions must slope in the same direction at any point of intersection. Thus, if the two patent policies are strategic complements in one country, they are strategic complements in both.

Returning to the case with $\beta \leq 0$, it is easy to show using (11) and $d\gamma/dT_i \leq 0$ that the $SS$ curve must have a slope that is everywhere greater in absolute value than $M_S/M_N$, while the $NN$ curve must have a slope that is everywhere smaller in absolute value than $M_S/M_N$.[15] It follows that $SS$ is steeper than $NN$ at any point of intersection of the two curves. This ensures stability of the policy setting game. It also guarantees uniqueness of the Nash equilibrium.

We can summarize the most important findings in this section as follows.

**Proposition 1** *Let the research technology be* $\phi_i = \left[b[L_{Ri}/a_i]^{\beta} + (1-b)H_i^{\beta}\right]^{1/\beta}$ *in country* $i$, *for* $i = S, N$. *If* $\beta \leq 0$, *then the two patent policies are strategic substitutes*

[15] We have not discussed the shape of the best response functions where they hit the axes or where the constraint that $T_i \leq \bar{T}$ begins to bind. The $SS$ curve becomes vertical if it hits the vertical axis at a point below $T_N = \bar{T}$. It also becomes vertical if the South's best response is $\bar{T}$ for some positive value of $T_N$. Similarly, the $NN$ curve becomes horizontal if either it hits the horizontal axis before $T_S = \bar{T}$ or if the North's best response is $\bar{T}$ for some positive value of $T_S$. Thus, the $SS$ curve must be steeper than the $NN$ curve at any point of intersection, even if these additional segments of the best response functions are taken into account.

*in both countries (i.e., the best response curves slope downward) and there exists a unique and stable Nash equilibrium of the policy setting game.*

# 4 Why are Patents Longer in the North?

Governments in the North typically grant longer patents and provide stronger patent protection more generally than their counterparts in the South.[16] In this section, we identify sufficient conditions under which patents in the North will be longer in duration than those in the South in the Nash equilibrium of a noncooperative policy game. Our goal here is to understand the reasons why the North may have a greater incentive to grant long patents than the South. We shall also examine how the equilibrium patent policies respond to changes in the endowments of human capital and to changes in the size of the market in each region.

We organize our discussion of the national differences in equilibrium policy choices around the following proposition.

**Proposition 2** *Suppose $M_N > M_S$ and $H_N > H_S$. Then $\tau_N \geq \tau_S$ in any Nash equilibrium of the patent policy game. Moreover, $\tau_N > \tau_S$ unless $\tau_S = \bar{\tau}$.*

The proposition is readily proved using the expressions for the best response functions in (11).[17] First, recall that with a CES research technology, $\mu_i = H_i/(H_S + H_N)$. Thus, $H_N > H_S$ implies $\mu_N > \mu_S$. The left-hand side of (11) is a decreasing function of $\mu_i$. If we cancel the terms that are common to the two best response functions, the remaining expression on the right-hand side is an increasing function of $M_i$ and a decreasing function of $T_i$. It follows that if $\mu_N > \mu_S$ and $M_N > M_S$, a pair of policies can be mutual responses only if $T_N > T_S$.

Our answer to the question in the section heading is that the North has a larger market for innovative goods and has a much greater capacity to conduct R&D. Why

[16] See, for example, Ginarte and Park (1997) who have constructed an index of patent rights and have shown that this index is highly correlated with per capita GDP.

[17] There are some details involving corner solutions with $\tau_S = 0$ or $\tau_N = \bar{\tau}$ that we leave to the interested reader.

do these characteristics induce the Northern government to grant longer patents in a noncooperative equilibrium than its counterpart in the South? The reasons are somewhat subtle.

Recall from the discussion in Section 2 that having a large market is not per se a reason for a government to grant long patents. The optimal patent length in a closed economy can in fact be independent of or even decreasing with market size, because both the marginal benefit of longer patents and the marginal costs of the associated distortions are proportional to $M$ for given $\gamma$, and the supply elasticity may remain the same or even decline as more resources are employed in R&D. The role of market size in generating different incentives for the governments has to do, instead, with the relative effectiveness of the two policy instruments. If $M_N$ is larger than $M_S$, innovative firms earn a majority of their profits in the North. Then, a given change in $T_N$ will generate a larger response of global innovation than would the same change in $T_S$. Since each policy generate deadweight loss in the country that affords the protection, the country that can more effectively stimulate innovation with a given lengthening of its patents will have an incentive to grant longer patents, all else equal.

The endowment of human capital proxies in our model for the capacity to conduct R&D. With $H_N > H_S$, a majority of the world's research is carried out in the North. As a consequence, a majority of the world's profits from innovative products accrue to residents of the North. In the North, the marginal cost of lengthening patents reflects the attendant loss in consumer surplus on all innovative products less the profits that are captured by Northern producers. Similarly, the marginal cost of lengthening patents in the South reflects the loss of consumer surplus there less the profits captured by Southern producers. But since the Northern producers earn a majority of the profits, the offset to marginal cost is larger in the North than in the South. Accordingly, the government of the North has less of a temptation to terminate its patents earlier than does the government of the South.

We turn next to the comparative static properties of the model. For this, we concentrate on the case in which the best response functions are downward sloping,

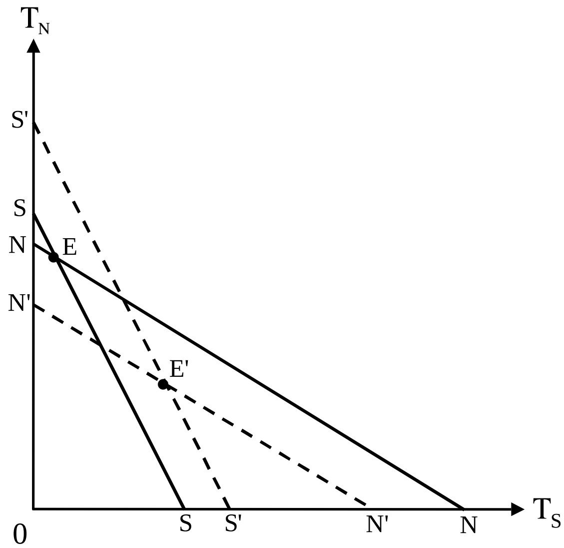


Figure 2: Comparative Static Effects of an Increase in $H_S/H_N$

which necessarily arises when (but not only when) $\beta \leq 0$.

Consider first the factor endowments. An equiproportionate change in $H_S$ and $H_N$ has no effect on $\mu_S$ or $\mu_N$, and thus no effect on the best response functions or the Nash equilibrium. Policy outcomes change only when there is a change in the *relative* endowments of human capital in the two countries. Suppose $H_S/H_N$ rises. This increases the share of innovation that occurs in the South ($\mu_S$) and reduces the share in the North ($\mu_N$). From (11) we see that the $SS$ curve shifts to the right while the $NN$ curve shifts downward; see Figure 2. The equilibrium shifts from $E$ to $E'$, with a reduction in patent duration in the North and an increase in patent duration in the South. This result is consistent with the Ginarte and Park (1997) finding that patent rights are positively correlated in a cross-national sample with secondary school enrollment rates and with the share of R&D in GDP.

We turn to the effects of market size. If $M_S$ and $M_N$ grow equiproportionately, the term $M_i T_i/(M_S T_S + M_N T_N)$ on the right-hand side of (11) is not affected at the initial values of $T_S$ and $T_N$. Then, if $\beta = 0$ (Cobb-Douglas research technology), $\gamma$ also is constant, and there is no effect on patent policy in either country. However, if $\beta < 0$, the extra resources that are drawn into R&D reduce the supply elasticity of innovation with respect to the value of a patent. Then the duration of patent protection falls in both countries. This is similar to our finding for a closed economy that, when $\beta < 0$, the optimal patent length shrinks when the market for differentiated products expands.

Next consider an expansion in the size of the Southern market, with no change in market size in the North. If $\beta = 0$, $\gamma$ is constant, and an increase in $M_S$ has qualitatively the same effects as an increase in $\mu_S$; these effects are shown in Figure 2, where we see that patent length in the South grows while that in the North shrinks. However, if $\beta < 0$, the increase in $M_S$ reduces $\gamma$ at the initial values of $T_S$ and $T_N$. Relative to the situation depicted in Figure 2, there is a further downward shift in $NN$ and an offsetting leftward shift in $SS$. Indeed, if the supply elasticity of innovation falls by enough, the $SS$ curve might even shift to its left relative to its initial location before the market expansion. In such circumstances $T_S$ might fall as $M_S$ grows.

# 5 International Patent Agreements

In this section, we study international patent agreements. We begin by characterizing the combinations of patent policies that are jointly efficient for the two countries.[18] Then we compare the Nash equilibrium outcomes with the efficient policies, to identify changes in the patent regime that ought to be effected by an international treaty. Finally, we address the issue of policy harmonization. By that point, we will have seen that harmonization is not necessary for global efficiency. We proceed to investigate the distributional properties of an agreement calling for harmonized patent policies and ask whether both countries would benefit from such an agreement in the absence of some form of direct compensation.

[18] Ours is a constrained efficiency, becuase we assme that innovation must be done privately, and that patents are the only policies available to encourage R&D. We do not, for example, allow the governments to introduce R&D subsidies, which if feasible, might allow them to achieve a given rate of innovation with shorter patents and less deadweight loss.

## 5.1 Efficient Patent Regimes

We shall begin by showing that the sum of the welfare levels of the two countries depends only on a measure $Q$ of the overall protection afforded by the international patent system. This means that the same aggregate world welfare level can be achieved with different combinations of $\tau_S$ and $\tau_N$ that imply the same overall level of protection. One particular level of $Q$ – call it $Q^*$ – maximizes the sum of the countries' welfare levels. For a wide range of distributions of world welfare, efficiency is achieved by setting the individual patent lengths so that the overall index of patent protection is $Q^*$.

In particular, let $Q = M_S T_S + M_N T_N$. This measure of global patent protection weights the discounted value of a one dollar flow extending for the duration of a patent in each country by the size of the country's market. A firm that earns a flow of profits $\pi$ per consumer for a period of length $\tau_S$ in the South and $\tau_N$ in the North earns a total discounted sum of profits equal to $Q\pi$. Thus, $Q$ governs the allocation of resources to R&D in each country, regardless of the particular combination of patent policies in the separate countries.

Consider the choice of patent policies $\tau_N$ and $\tau_S$ that will take effect at time 0 and apply to goods invented thereafter. The expressions for the countries' gross welfare levels at time 0 are analogous to that for a closed economy, as recorded in equation (6). The aggregate welfare in country $i$, discounted to time 0, is given by

$$U_i(0) = \Lambda_{i0} + \frac{w_i(L_i - L_{Ri})}{\rho} + \frac{M_i(\phi_S + \phi_N)}{\rho}\left[T_i C_m + (\bar{T} - T_i)C_c\right] + \frac{\phi_i}{\rho}\pi\left(M_S T_S + M_N T_N\right), \text{ for } i = S, N, \tag{12}$$

where $\Lambda_{i0}$ is the fixed amount of discounted surplus that consumers in country $i$ derive from goods that were invented before time 0.

Summing the expressions in (12) for $i = S$ and $i = N$, we find that

$$\rho\left[U_S(0) + U_N(0)\right] = \rho\left(\Lambda_{S0} + \Lambda_{N0}\right) + w_S(L_S - L_{RS}) + w_N(L_N - L_{RN}) + \left(M_S + M_N\right)\bar{T}(\phi_S + \phi_N)C_c - Q\left(\phi_S + \phi_N\right)\left(C_c - C_m - \pi\right) \tag{13}$$

Since $v_S = v_N = \pi Q$, $L_{RS}$ and $L_{RN}$ are functions of $Q$.[19] The same is true of $\phi_S$ and $\phi_N$. It follows that different combinations of $\tau_S$ and $\tau_N$ that yield the same value of $Q$ also yield the same level of aggregate world welfare.[20]

If international transfer payments are feasible, then a globally efficient patent regime must have $M_S T_S + M_N T_N = Q^*$, where $Q^*$ is the value of $Q$ that maximizes the right-hand side of (13).[21] Notice that a range of efficient outcomes can be achieved without the need for any international transfers. By appropriate choice of $\tau_N$ and $\tau_S$, the countries can be given any welfare levels on the efficiency frontier between that which they would achieve if $T_S = 0$ and $T_N = Q^*/M_N$ and that which they would achieve if $T_S = Q^*/M_S$ and $T_N = 0$.[22]

In Figure 3, the bold curve depicts the combinations of $T_S$ and $T_N$ that may be chosen in an efficient world patent regime when international transfer payments are not feasible. The welfare of the South increases, and that of the North decreases, as we move down the vertical segment from $(0, \bar{T})$, down the downward-sloping line between $(0, Q^*/M_N)$ and $(Q^*/M_S, 0)$, and finally to the right along the horizontal segment joining $(Q^*/M_S, 0)$ and $(\bar{T}, 0)$. Along the downward-sloping segment, aggregate world welfare is constant. Aggregate welfare declines as we move up along the vertical segment or to the right along the horizontal segment; this deadweight loss is the cost

[19] In country $i$, the allocation of labor to research is determined by

$$\pi Q F_L(L_{Ri}/a_i, H_i) = 1/a_i.$$

[20] This result is anticipated by a similar one in McCalman (1997), who studied efficient patent agreements in a partial equilibrium model of cost-reducing innovation by a single, global monopolist.

[21] The first-order condition for maximizing $\rho[U_S(0) + U_N(0)]$ implies

$$C_c - C_m - \pi = \gamma \left\{ C_m + C_c \left[ \frac{(M_S + M_N)\bar{T} - Q^*}{Q^*} \right] \right\}.$$

The second-order condition is satisfied at $Q = Q^*$ when $\beta \leq 1/2$.

[22] This statement ignores the ceiling on patent lengths imposed by the finite economic life of differentiated products. A more precise statement is that a range of distributions of maximal world welfare can be achieved by varying $T_S$ between $T_S = \max\{0, (Q^* - M_N\bar{T})/M_S\}$ and $\min\{Q^*/M_S, \bar{T}\}$ while varying $T_N$ between $T_N = \min\{Q^*/M_N, \bar{T}\}$ and $\max\{0, (Q^* - M_S\bar{T})/M_N\}$ in such a way that $M_S T_S + M_N T_N = Q^*$.

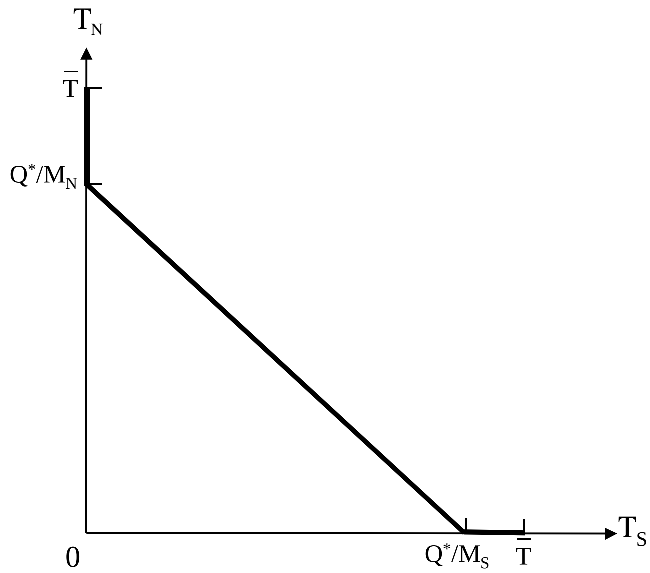


Figure 3: Efficient Patent Regimes

of achieving such skewed distributions of world welfare in the absence of international transfers.

## 5.2 Pareto-Improving Patent Agreements

How do the efficient combinations of patent lengths compare to the policies that emerge in a noncooperative equilibrium? The answer to this question — which informs us about the likely features of a negotiated patent agreement — is illustrated in Figure 4. The figure depicts the best response functions and the efficient policy combinations on the same diagram.

We show the $QQ$ line being situated to the right of the $SS$ curve and above the $NN$ curve. This is a general feature of our model, not dependent on any assumptions about the countries' research technologies. The reasons are clear. Starting from a point on the South's best response function, a marginal increase in the length of patents in the South must increase world welfare. Such a lengthening of Southern patents has only a second-order effect on welfare in the South, but it conveys two positive externalities to the North. First, a lengthening of Southern patents provides

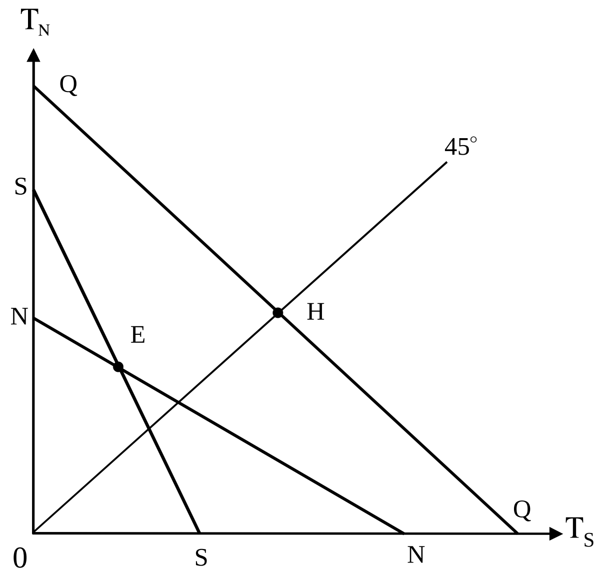


Figure 4: Comparison of Nash Equilibrium and an Efficient Patent Regime

extra monopoly profits to Northern innovators, which contributes to aggregate income there. Second, an increase in $\tau_S$ enhances the incentives for R&D, inducing an increase in both $\phi_S$ and $\phi_N$. The extra product diversity that results from this R&D creates additional surplus for Northern consumers.[23]

By the same token, a marginal increase in the length of Northern patents from a point along $NN$ increases world welfare. Such a change in policy enhances profit income for Southern firms, and encourages additional innovation in both countries. It follows, of course, that the $QQ$ line must lie outside the Nash equilibrium. We record our finding in

**Proposition 3** *Let* $(T_S, T_N)$ *be an interior equilibrium in the noncooperative policy game and let* $(T_S^*, T_N^*)$ *be any efficient combination of patent policies. Then* $M_S T_S^* + M_N T_N^* > M_S T_S + M_N T_N$.

The proposition implies that, starting from any interior Nash equilibrium, an efficient

[23] A more formal proof that the $QQ$ line lies outside the $SS$ curve and the $NN$ curve is available from the authors upon request.

patent treaty must lengthen patents in at least one country. It also implies that the treaty will strengthen global incentives for R&D and induce more rapid innovation in both countries.

## 5.3 Harmonization

Commentators sometimes claim that it would be desirable to have universal standards for intellectual property protection and for many other national policies that affect international competition. The arguments for harmonization are not always clear, but they seem to be based on a desire for global efficiency. Yet it is hardly obvious why efficiency should require identical policies in countries at different stages of economic development. In this section, we examine the aggregate and distributional effects of international harmonization of patent policies.

As should be apparent from the preceding discussion, harmonization of patent policies is neither necessary nor sufficient for global efficiency, regardless of whether international transfer payments are feasible or not. A regime of harmonized policies will only be efficient if the common duration of patents in the two countries is such that $Q = Q^*$. And any combination of patent policies that provides the proper global incentives for R&D will be efficient, no matter whether the patent lengths in the two countries are the same or not.

If patents are longer in the North than in the South in an initial Nash equilibrium, then harmonization might be achieved either by a unilateral lengthening of patents in the South or by a combination of policy changes in the two countries. A unilateral lengthening of Southern patents is bound to harm the South (absent any side payments), because the equilibrium $\tau_S$ is a best response by the South and any unilateral deviation from a country's best response is, by definition, damaging to its interests.[24] As for harmonization that might be achieved through a combination of

[24]See also Lai and Qiu (2002), who consider the welfare effects of harmonizing IPR protection at the standard that would be chosen by the North in a non-cooperative equilibrium. In a model of once-off investment in R&D, they show that such a change in the South's policy from the Nash equilibrium level would benefit the North by more than it would harm the South.

policy changes, we focus on a treaty that would achieve global efficiency. Such a treaty is represented by point $H$ in Figure 4. Efficient harmonization surely requires an increase in patent duration in the South, since $\tau_N > \tau_S$ at $E$ and $QQ$ lies outside this point. If $\beta \leq 0$, it also requires an increase in patent length in the North.[25] If $M_N \geq M_S$ and $H_N \geq H_S$, the North definitely gains from efficient harmonization.[26] However, the South may be worse off at point $H$ than in the Nash equilibrium at point $E$, unless some form of compensation is provided by the North. In general, the larger are $M_N/M_S$ and $H_N/H_S$, the more likely it is that the South would lose from efficient harmonization.

Summarizing, we have

**Proposition 4** *Suppose $M_N \geq M_S$, $H_N \geq H_S$, and $\beta \leq 0$. Then efficient harmonization requires a lengthening of patents in both countries. The North necessarily gains from efficient harmonization, while the South may gain or lose.*

We conclude that harmonization has more to do with distribution than with efficiency, and that incorporation of such provisions in a treaty like the TRIPs agreement might well benefit the North at the expense of the South.[27]

[25] First, we note that when $\beta \leq 0$, point $H$ lies above the intersection of the $NN$ curve with the vertical axis. This can be seen by substituting $T_N = T_S$ in the first-order condition for maximizing $\rho\,[U_S(0) + U_N(0)]$ and comparing the resulting expression for $T_N = Q^*/(M_N + M_S)$ with the expression for $T_N$ that comes from (11) when $T_S = 0$. Then, since the $NN$ curve is downward sloping when $\beta \leq 0$, the fact that it starts below point $H$ implies that the Northern patent length is longer at point $H$ than it is at point $E$.

[26] If $M_N \geq M_S$ and $H_N \geq H_S$, the common patent length that maximizes the welfare of the North is greater than the common patent length that maximizes aggregate world welfare. Therefore, the North gains from a unilateral increase in $\tau_S$ that brings the Southern patent policy into conformity with the Nash equilibrium policy in the North, and further gains from an increase in the common policy until $Q = Q^*$.

[27] McCalman (2000) estimates the income transfers implicit in the TRIPs agreements and finds that international patent harmonization benefits the United States at the expense of the developing countries as well as Canada, the United Kingdom and Japan.

# 6 Patent Policy with Many Countries

In this section, we extend our analysis to a trading world with many countries. Our main finding is that adding countries exacerbates the free-rider problem that plagues the noncooperative policy equilibrium. Small countries are inclined to allow others to provide the incentives for innovation so as to avoid the deadweight losses in their home markets. In the limit, as the number of countries grows large and each one is small in relation to the world economy, the unique Nash equilibrium has universal patents of length zero. Then, a patent treaty is critical for creating incentives for private innovation.

We assume that there are $J$ countries, and that country $i$ has market size $M_i$, human capital endowment $H_i$, and labor productivity $1/a_i$. The research technology in country $i$ is $\phi_i = F(H_i, L_{Ri}/a_i) = \left[b[L_{Ri}/a_i]^\beta + (1-b)H_i^\beta\right]^{1/\beta}$, with $\beta \leq 1/2$. All consumers share the preferences given in (8).

Suppose that there is no cooperation between nations in setting their patent policies. In country $i$, either $T_i = 0$ and the marginal cost of providing the first bit of patent protection exceeds the marginal benefit, or $T_i = \bar{T}$ and the marginal benefit of providing the last bit of patent protection exceeds the marginal cost, or $0 < T_i < \bar{T}$ and the marginal benefit of lengthening patents equals the marginal cost. Equality between marginal benefit and marginal cost implies

$$C_c - C_m - \mu_i \pi = \frac{M_i}{Q}\gamma[T_i C_m + C_c(\bar{T} - T_i)] \ , \tag{14}$$

where $Q = \sum_j M_j T_j$ measures the global patent protection in the Nash equilibrium.

Observe first that as $\mu_i \to 0$, the left-hand side of (14) approaches $C_c - C_m$; a small country captures virtually none of the monopoly profits from innovative products, so the cost of a patent per consumer and product is the difference between the competitive and monopoly levels of consumer surplus. But as $M_i \to 0$, the right-hand side of (14) approaches zero, because a small country provides innovators with virtually none of their global profits and so worldwide innovation is hardly responsive to a change in such a country's patent policy. It follows that a small country (in a world with some large countries) will set its patent length to zero in a Nash equilibrium.

If all countries choose positive patent lengths that are less than $\bar{T}$, equation (14) holds for every $i$. Then we can sum (14) across the $J$ countries, which gives

$$J\left(C_c - C_m\right) - \pi = \gamma\left[C_m - C_c + \frac{C_c\left(\sum_j M_j\right)\bar{T}}{Q}\right]. \qquad (15)$$

Then, for a given size of the world market, $Q$ depends only on the *number* of countries $J$ and not on the distribution of consumers and human capital across countries. Moreover, if $\beta \leq 0$, $Q$ is a declining function of $J$; the greater is the number of countries, the weaker are the global incentives for innovation in a noncooperative equilibrium. As the number of countries grows large (holding constant the size of the world market), the aggregate incentives for innovation approach zero.[28] Evidently, the free-rider problem becomes increasingly severe as the number of independent decision makers in the world economy expands.

Finally, note that the requirements for global efficiency do not depend on the number of countries. Again, the sum of all national welfare levels is a function of the aggregate world incentive for innovation. This sum is maximized when

$$C_c - C_m - \pi = \gamma\left[C_m - C_c + \frac{C_c\left(\sum_j M_j\right)\bar{T}}{Q^*}\right]. \qquad (16)$$

Thus, if international compensation is possible, an efficient global patent treaty will have $\sum_j M_j T_j = Q^*$, where $Q^*$ is solved from (16). Notice that $Q^*$ must exceed $Q$, the aggregate patent protection in the Nash equilibrium. Even if international compensation is not feasible, an efficient agreement will have $\sum_j M_j T_j = Q^*$ for a range of distributions of world welfare.

# 7 Conclusions

We have developed a simple model of endogenous innovation and have used it to study the incentives that governments face in choosing their patent policies. Our

[28]Suppose $Q$ were to approach a finite number as $J \to \infty$. Then $\gamma$ would approach a finite number as well, and the right-hand side of (15) would be finite. But the left-hand side of (15) approaches infinity as $J \to \infty$.

model features a familiar trade-off between the static benefits of competitive pricing and the dynamic benefits of increased innovation. For a closed economy, we derived a simple formula for the optimal patent length that relates the deadweight loss induced by a marginal lengthening of the period of patent protection to the surplus that results from the extra innovation.

In an open economy, differences in market size and differences in capacity for R&D generate national differences in optimal patent policies. We focused on policies that are applied with *national treatment*; that is, regimes that require equal protection for foreign and domestic applicants. A country's optimal patent duration is found by equating the sum of the extra deadweight loss that results from lengthening the patents granted to domestic firms and the extra surplus loss that results from extending the monopoly pricing by foreign firms with the benefits that flow from providing greater incentives for innovation to firms worldwide. A country's optimal patent length depends on the policies set by its trading partner, because the strength of foreign patent rights affects the responsiveness of global innovation to a change in a country's own patent duration.

We found that having a larger market for innovative products typically enhances a government's incentive to grant longer patents. Also, a government's relative incentive to provide patent protection typically increases with its relative endowment of human capital. In a noncooperative equilibrium, patent duration will be longer in the North than in the South if the North has a larger market for innovative products and a greater capacity for R&D.

Starting from a Nash equilibrium, countries can benefit from negotiating an international patent agreement. A treaty can ensure that national policies reflect the positive externalities that flow to foreign residents when a country extends the length of its patents. To achieve (constrained) efficiency, an international agreement must strengthen aggregate world patent protection relative to the Nash equilibrium. Harmonization of patent policies is neither necessary nor sufficient for the efficiency of the global patent regime. If patent policies are harmonized at an efficient level, the move from a Nash equilibrium typically will benefit the North but possibly harm the

South.

Our conclusions are essentially the same for a world with more than two countries. Countries with larger markets and more human capital will provide longer patents in a noncooperative equilibrium than those with smaller markets and less human capital. Indeed, a country that is small in relation to the world economy has no incentive whatsoever to grant patents. The greater is the number of independent countries, the more severe is the free-rider problem inherent in the setting of national patent policies. Thus, the value of an international patent agreement grows with the number of independent sovereign decision makers.

Our analysis can be extended to more general environments. For example, in an earlier version of this paper (Grossman and Lai, 2002), we allowed for cross-national differences in relative labor productivity in the two industries. With comparative advantage in production, the productivity gap in the industry that produces differentiated products may not be offset by the gap in relative wages. Then the production costs for innovative products will be higher in one region or the other. This can create an asymmetry in the life cycle of a new good depending upon whether patents are longer in the North or in the South. We showed how such an asymmetry may generate multiple equilibria in the policy game.

Another possible extension would allow different preferences in different countries. With different demands, the marginal cost of lengthening patent protection will vary around the globe. Then differences in the elasticities of demand for innovative products will be another factor that affects the governments' relative incentives for granting long patents. Moreover, asymmetries in demand would be reflected in the characteristics of a globally efficient patent regime. An efficient regime would equalize across countries the marginal deadweight loss associated with providing a given push to global innovation. Efficiency requires longer patents in countries that have more inelastic demands for innovative products, all else the same.

# References


[1] Chin, Judith and Grossman, Gene M. (1990), "Intellectual Property Rights and North-South Trade," in R.W. Jones and A.O. Krueger, eds., *The Political Economy of International Trade*, Cambridge MA: Basil Blackwell Publishers.

[2] Deardorff, Alan V. (1992), "Welfare Effects of Global Patent Protection," *Economica* 59, 35-51.

[3] Ginarte, Juan C. and Park, Walter G. (1997), "Determinants of Patent Rights: A Cross-National Study," *Research Policy* 26, 283-301.

[4] Grossman, Gene M. and Lai, Edwin L.-C. (2002), "International Protection of Intellectual Property," National Bureau of Economic Research Working Paper No. 8794, Cambridge MA.

[5] Johnson, Harry G. (1953-54), "Optimal Tariffs and Retaliation," *Review of Economic Studies* 21, 142-153.

[6] Lai, Edwin L.-C. and Qiu, Larry D. (2002), The North's Intellectual Property Rights Standard for the South?" *Journal of International Economics*, forthcoming.

[7] Maskus, Keith E. (2000a), *Intellectual Property Rights in the Global Economy,* Washington D.C.: The Institute for International Economics.

[8] Maskus, Keith E. (2000b), "Parallel Imports," *The World Economy* 23, 1269-1284

[9] Mayer, Wolfgang (1981), "Theoretical Considerations on Negotiated Tariff Adjustments," *Oxford Economic Papers* 33, 135-153.

[10] McCalman, Phillip (1997), "National Patents, Innovation and International Agreements," University of California at Santa Cruz, mimeo.

[11] McCalman, Phillip (2001), "Reaping What you Sow: An Empirical Analysis of International Patent Harmonization," *Journal of International Economics* 55, 161-185.

[12] Nordhaus., William D. (1969), *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change*, Cambridge MA: The M.I.T. Press.

[13] Scotchmer, Suzanne (2001), "The Political Economy of Intellectual Property Treaties," Institute of Business and Economic Research, Working Paper No. E01-305, University of California at Berkeley.

# 8 Appendix

In this appendix we show that, for a closed economy, when $T$ solves (7) and $\beta \leq 1/2$, the second-order condition for an optimal patent length is satisfied. Similar calculations ensure that $\beta \leq 1/2$ is sufficient for the second-order condition to be satisfied for the best response given by (11) for an open economy.

Let us rewrite the first-order condition (7) as

$$\frac{\gamma}{T}\left\{C_c\bar{T} - T\left[(C_c - C_m) + (C_c - C_m - \pi)\frac{1}{\gamma}\right]\right\} = 0.$$

Since $\gamma/T > 0$, the term in curly brackets must vanish at any local extremum. We will show that at any such point the term in curly brackets is a decreasing function of $T$; i.e., that

$$-\left[(C_c - C_m) + (C_c - C_m - \pi)\frac{1}{\gamma} - T\frac{(C_c - C_m - \pi)}{\gamma^2}\frac{d\gamma(T)}{dT}\right] < 0. \tag{17}$$

This means that any point satisfying the first-order condition is a local welfare maximum. Since the welfare function is continuous and differentiable, it follows that there can be at most one local extremum point, and that the value of $T$ that generates this point is the unique welfare-maximizing patent duration.

It is straightforward to calculate that

$$\gamma \equiv -\left[\frac{(F_L)^2}{FF_{LL}}\right] = \frac{b}{(1-b)(1-\beta)}\left(\frac{L_R}{aH}\right)^{\beta}$$

for the CES research technology. Meanwhile, $vF_L = w = 1/a$ and $v = M\pi T$ imply that $F_L = 1/aM\pi T$, or that

$$\frac{b}{a}\left[b + (1-b)\left(\frac{aH}{L_R}\right)^{\beta}\right]^{\frac{1-\beta}{\beta}} = \frac{1}{aM\pi T}$$

in the CES case. Using these two equations, we can express $\gamma$ as a function of $T$; we find that $\gamma = \left(\frac{b}{1-\beta}\right)\left[(bM\pi T)^{\frac{-\beta}{1-\beta}} - b\right]^{-1}$.

We can now compute $d\gamma(T)/dT$, and substitute the resulting expression into the left-hand side of (17), which then becomes

$$\Omega = -\left[C_c - C_m - (1-\beta)(C_c - C_m - \pi) + (C_c - C_m - \pi)(bM\pi T)^{\frac{-\beta}{1-\beta}}\left(\frac{1-2\beta}{b}\right)\right].$$

But $\beta \leq 1/2$ ensures that

$$
\begin{aligned}
\Omega &\leq -\left[C_c - C_m - (1-\beta)(C_c - C_m - \pi) + (1-2\beta)\,(C_c - C_m - \pi)\ \right] \\
&= -\left[C_c - C_m - \beta(C_c - C_m - \pi)\right] \\
&< 0\ .
\end{aligned}
$$

So the second-order condition is satisfied when $\beta \leq 1/2$ and the $T$ that solves (7) – if it exists – is the unique welfare-maximizing patent policy.