NBER WORKING PAPER SERIES

INTERNATIONAL PROTECTION OF INTELLECTUAL PROPERTY

Gene M. Grossman
Edwin L.-C. Lai

Working Paper 8704
http://www.nber.org/papers/w8704

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2002

The authors thank Elhanan Helpman, Giovanni Maggi, Keith Maskus and Lars Svenssion 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. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

International Protection of Intellectual Property
Gene M. Grossman and Edwin L.-C. Lai
NBER Working Paper No. 8704
January 2002
JEL No. O34, F13

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, in their capacities for innovation, and in their absolute and comparative advantage in manufacturing. 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.

Gene Grossman
International Economics Section
Department of Economics
Princeton University
Princeton, NJ 08544
and NBER
grossman@princeton.edu

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

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, in their capacities for innovation, and in their absolute and comparative advantages in manufacturing. 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

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

Nordhaus. We derive a neat formula that characterizes the optimal patent policy in a closed economy, and discuss the determinants of the optimal patent length. One interesting finding is that the optimal duration of patents is independent of 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. The wage gap between countries creates a difference in manufacturing costs, which bears on the country's non-cooperative policy choices. In particular, we identify important differences in the incentives that the countries have to extend their patent lengths when the high wage country has the longer patents as compared to when the low wage country has the longer patents.

In Section 4, we ask, Why are patents longer in the North? We address this question by considering the comparative statics of the Nash equilibrium. The fact that the North has greater research capability than the South is not enough to explain its longer patents. But we are able to show that the Nash equilibrium patents will be longer in the North than in the South if the North's market for innovative products is at least as large as that in the South and if the North's relative advantage in R&D is sufficiently great.

We study international patent agreements in Section 5. First we derive the properties of an efficient global regime of patent protection. If the countries can make international transfer payments or compensate one another in policy areas other than IPR protection, then 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. Even if international transfers are not possible, the efficient policies will be ones that provide the optimal aggregate incentive for R&D, at least for a range of distributions of national welfare levels. However, when international transfers do not occur, the welfare levels of the North and the South will depend on which pair of efficient patent lengths is selected. 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 typically serves the interests of 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. Our findings concerning the non-cooperative and cooperative outcomes are summarized in Section 6.

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 t. 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 ρ, from the condition for intertemporal optimization.

[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.

Manufacturing requires only labor. Any firm can produce good y with b units of labor per unit of output. All known varieties of the differentiated product can 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 τ. We assume that patents are perfectly enforced.

The design of new varieties requires both labor and human capital. We take $\phi(z) = F[H, L_R(z)]$, where $\phi(z)$ is the flow of new inventions at time z, H is the (constant) stock of human capital, and $L_R(z)$ is the amount of labor devoted to R&D. 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 assume that $F(\cdot)$ is homogeneous of degree one and that $\gamma(L) \equiv -\left(F_L\right)^2 / \left(FF''\right)$ is a non-increasing function of L. We shall see that γ is the elasticity of research output with respect to the value of a patent and that our restriction that it is non-increasing ensures that the second-order condition for the optimal patent length is satisfied. If the research technology is Cobb-Douglas — an example that we shall use in several places — then γ is constant and equal to the ratio of the cost share of labor to the cost share of human capital.[3] If the research technology has a constant elasticity of substitution, γ is a non-increasing function of L for any elasticity of substitution less than or equal to one.

We describe now the static and dynamic equilibrium for an economy that has a patent duration of τ. 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 π 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.

[3] In other words, if $F(H, L_R) = H^{1-\beta} L_R^{\beta}$, then $\gamma = \beta/(1-\beta)$.

This pricing of the good continues until the good becomes obsolete. Meanwhile, the homogeneous good always carries the competitive price of bw, which, because this good is the numeraire, implies that $w = 1/b$. 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,

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

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.[4] 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 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 τ, 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 τ 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

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

where Λ_0 is the discounted present value of the consumer surplus and profits derived from goods invented before time 0, and where $T \equiv (1 - e^{-\rho\tau})\,/\rho$ and $\bar{T} \equiv (1 - e^{-\rho\bar{\tau}})\,/\rho$. Note that T is the present discounted value of a flow of one dollar from time 0 to time τ, 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 τ, after recalling that $\phi = F(H, L_R)$ and that L_R is a function of τ via (3) and (4).[5] It is more intuitive, however, to describe

[4] 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.

[5] Equivalentlly, we can maximize $\rho U(0)$ over the choice of T. Note that C_m, C_c and π 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

the social costs and benefits that derive from extending the patent length marginally from a given length τ. The cost of lengthening the period of patent protection is that the economy suffers the deadweight loss of $M\left(C_c - C_m - \pi\right)$ 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 τ, and continues thereafter. The flow of new products is $\dot{n}$ per unit time. Thus, 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 can calculate that

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

thus, γ is the elasticity of innovation with respect to the value of a patent, as previously claimed. Also, (4) implies

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

into (6) and rearranging terms, we can write the maximand as

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

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.

We substitute these terms into the expression for marginal benefit, and equate this to the marginal cost. This gives an implicit formula for the optimal patent length, namely

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

The assumption that $\gamma' < 0$ ensures that the second-order condition is satisfied.

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 ρ), and the greater is the ratio of consumer surplus plus profits under monopoly to consumer surplus with competition. All of these findings accord well with intuition. One noteworthy feature of (7) is that the size of the market has no bearing on the optimal patent length in a closed economy. In a closed economy, the first-best level of R&D — that which maximizes discounted utility when all goods are competitively priced — 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. An increase in M enhances both the marginal benefit of extending patents and the marginal cost of doing so, and does so in equal proportions. Thus, the optimal patent length in a closed economy is invariant to market size.

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

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

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.

[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.

[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.

In country j, it takes b_j units of labor to produce one unit of the homogeneous good. We assume that $b_N < b_S$, and that, in equilibrium, the numeraire good is produced in both countries. This implies $w_N > w_S$. The technology for producing a differentiated product reflects the country in which the good was invented. A variety that was developed in country j requires a_j units of labor per unit of production, where $a_N \leq a_S$.[8] Of course, the rights to produce such goods may be limited by patent protection. For now, we rule out direct foreign investment, so the proprietary owner of a technology for producing a differentiated product must undertake its manufacturing in the same country in which its R&D was conducted.

New goods are invented in each region according to $\phi_j = F(H_j, L_{Rj})$, where H_j is the human capital endowment of country j, and L_{Rj} is the labor devoted to R&D in country j. The natural case to consider is $H_N > H_S$, but we do not impose this as a restriction.

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 τ_j to all inventors of differentiated products regardless of there 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 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

[8] We could have specified, alternatively, that the technology for producing a differentiated product depends on where the good is produced, not where it was invented. The analysis would not be much different.

[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.

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 good 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 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 τ_N and τ_S, the life cycle of a typical differentiated product invented in the North and that of a typical product invented in the South. For this, we must distinguish a global patent regime with $\tau_N > \tau_S$ from a regime with $\tau_N < \tau_S$. Consider first a product invented in the North. During an initial phase after the product is developed, the inventor holds a patent in both countries.

[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.

Then the patent holder earns a flow of profits $M_N \pi_N$ from sales to consumers in the North and $M_S \pi_N$ from sales to consumers in the South, where π_N is earnings per consumer for a monopoly selling a typical Northern brand. Households in the North realize a flow of consumer surplus of $M_N C_m^N$ from these purchases, while their those in the South realize a flow of surplus of $M_S C_m^N$, where C_m^j is the surplus enjoyed by a typical consumer of a monopolistically-priced good produced in country j. If $\tau_N > \tau_S$, the inventor's patent will expire first in the South. At that time, the good will be imitated by competitive firms, which will sell it there for a price of $w_S a_N$. The patent holder ceases to realize any profits from sales in the South, but continues to earn profits in the North. The flow of consumer surplus in the South rises to $M_S C_c^{NS}$, where C_c^{jk} denotes the consumer surplus generated per consumer by a product that was invented in country j and is sold at a competitive price by producers in country k. Consumer surplus in the North continues to be $M_N C_m^N$, until the patent expires there. Then imitation by low-cost Southern producers causes Northern consumer surplus to rise to $M_N C_c^{NS}$, while the Northern inventor loses his remaining source of income. At a time $\bar{\tau}$ from the moment of invention, the good becomes obsolete and all flows of consumer surplus cease.

If, instead, $\tau_N < \tau_S$, the Northern patent will expire first. Then the good can be imitated by firms in the North for sale to Northern consumers, but the ongoing patent protection in the South prevents both imports and production there. For a while, the consumers in the North enjoy (in the aggregate) an intermediate flow of consumer surplus of $M_N C_c^{NN}$, because the competitively-priced good is produced by firms in the high-wage country. The inventor loses all of its earning capacity in the North as soon as this imitation takes place. When a period τ_S has elapsed from the time of invention, competitive production begins in the South, and these low-cost producers capture both markets. Then consumer surplus in the North rises (again) to $M_N C_c^{NS}$, consumer surplus in the South rises to $M_S C_c^{NS}$, and the inventor loses his remaining source of profits.

For a good invented in the South, the life cycle is similar, but with one small difference. If $\tau_N > \tau_S$, the consumer surplus flows are first $M_S C_m^S$ and $M_N C_m^N$

during a period of patent protection in both countries, then $M_S C_c^{SS}$ for consumers in the South after the patent protection has expired there, and finally $M_N C_c^{SS}$ for consumers in the North, once its longer patent expires. The inventor earns a flow of profits of $(M_S + M_N)\pi_S$ during the initial phase with protection in both markets, $M_N \pi_S$ during the second phase with protection only in the North, and nothing during the final phase with a competitive world market. However, if $\tau_N < \tau_S$, the loss of patent protection in the North does not spell an end to the inventor's profits there. This is because the Southern inventor can produce the good at a cost of $w_S a_S$, which is less than the cost facing a potential imitator in the North. If a Northern firm's per unit cost, $w_N a_S$, exceeds the price charged by a Southern monopoly while it holds full patent protection, then the patent holder's profits and the flow of consumer surplus in the North continue unabated after the expiration of the patent in the North. If the Northern unit cost falls short of the Southern monopoly price, then at the time when its Northern patent expires the Southern firm must shave its price in the North to a bit below $w_N a_S$, suffering in consequence a reduction in its profit flow to $M_N \tilde{\pi}_S + M_S \pi_S$, where $\tilde{\pi}_S$ is the profits per consumer of a Southern monopoly that faces potential competition from Northern imitators. The Northern consumers, in turn, realize a jump in their surplus flow to $M_N \tilde{C}_m^S$, where $\tilde{C}_m^S = \max\{C_c^{SN}, C_m^S\}$ is the per-consumer surplus in the North when a Southern monopolist faces potential competition from producers in the North. When the Southern patent eventually expires, the competitive producers in the South take over both markets, and the surplus per consumer rises everywhere to C_c^{SS}. At this point, the inventor ceases to capture any profits.

3.2 The Best Response Functions

We are now ready to derive the best response function for each country. The best response function for the South gives the patent length τ_S that maximizes aggregate welfare in the South as a function of a given τ_N. Similarly, the best response function for the North gives the patent length τ_N that maximizes Northern welfare, given the patent policy of the South. Conceptually, we proceed as follows. First, we examine the first-order condition that must be satisfied by τ_S if $\tau_S > 0$ and we impose the restriction that $\tau_N > \tau_S$. Then, we examine the first-order condition that must be satisfied if $\tau_S < \bar{\tau}$ and we suppose that $\tau_N < \tau_S$. Finally, we consider the South's optimal choice of patent length for given τ_N without any restrictions.

Given $\tau_N > \tau_S$, the South bears two costs from prolonging its patents slightly. First, this extends the period during which it suffers a static deadweight loss of $C_c^{SS} - C_m^S - \pi_S$ on goods invented in the South. Second, it prolongs the period during which its consumers realize surplus of only C_m^N instead of C_c^{NS} on goods that were 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 residents of the North. The marginal benefit to the South that comes from prolonging its patents reflects the increased incentive that Northern and Southern firms have to undertake R&D. If the welfare-maximizing τ_S is positive and less than τ_N, then the marginal cost of increasing τ_S must equal the marginal benefit, which implies

$$\begin{aligned}
&\phi_S(C_c^{SS} - C_m^S - \pi_S) + \phi_N(C_c^{NS} - C_m^N) \\
&\quad = \frac{\phi_S}{v_S}\gamma_S\pi_S\left[M_S C_m^S T_S + C_c^{SS}(\bar{T} - T_S)\right] + \frac{\phi_N}{v_N}\gamma_N\pi_N\left[M_S C_m^N T_S + C_c^{NS}(\bar{T} - T_S)\right],
\end{aligned}$$

where $\gamma_j = -(F_j'^2)/(F_j'' F_j)$, $T_j = \left(1 - e^{-\rho\tau_j}\right)/\rho$, and v_j is the value of a new patent for a good invented in country j. Noting that $v_j = \left(M_S T_S + M_N T_N\right)\pi_j$, this can be written as

$$\begin{aligned}
&\left(C_c^{SS} - C_m^S - \pi_S\right) + \omega(C_c^{NS} - C_m^N) \\
&\quad = \gamma_S\Omega_S\left[C_m^S + C_c^{SS}\left(\frac{\bar{T} - T_S}{T_S}\right)\right] + \omega\gamma_N\Omega_S\left[C_m^N + C_c^{NS}\left(\frac{\bar{T} - T_S}{T_S}\right)\right], \qquad (9)
\end{aligned}$$

where $\omega = \phi_N/\phi_S$ and $\Omega_j = M_j T_j/\left(M_S T_S + M_N T_N\right)$. The terms on the left-hand side of (9) reflect respectively the marginal cost of extending the deadweight loss in the South on Southern products and the marginal cost of postponing the competitive pricing of Northern products. The terms on the right-hand side of (9) reflect the benefit to Southern consumers from the induced spur to innovation in the South and North, respectively.

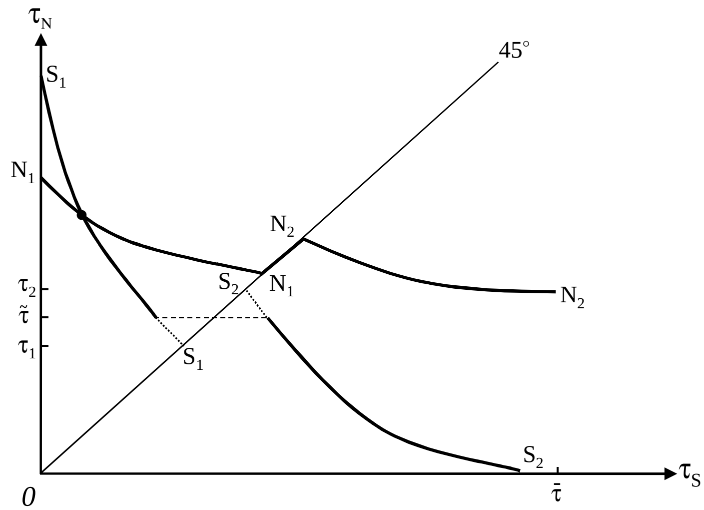

Figure 1: Nash Equilibrium with Continuous Best Response in North

Equation (9) is depicted by the curve S_1S_1 in Figure 1, including both the boldface and dotted portions (more on this in a moment). Notice that we have drawn the curve only for points above the 45° line, inasmuch as the stated first-order condition applies only when $\tau_N > \tau_S$. We illustrate a case in which the curve is downward sloping, as it must be when the countries have identical Cobb-Douglas research production functions.[11] Intuitively, as τ_N increases, labor is drawn into R&D in both countries. This reduces the responsiveness of global innovation to a given change in τ_S. If the fraction of world innovation that takes place in each country does not change by too much (as will be the case if γ_S is approximately equal to γ_N) then the South's best response is to set a shorter patent of its own.

Now suppose that $\tau_S > \tau_N$. In this case, the marginal cost to the South of prolonging its patents has an additional, negative term. It includes as before terms reflecting an extended period of deadweight loss on goods invented in the South and an extended period of diminished consumer surplus on goods invented in the North.

[11] With identical Cobb-Douglas research functions, γ_S and γ_N are identical constants, and ω is independent of τ_S and τ_N. Then the marginal benefit declines with τ_N and the second-order condition (which is satisfied) ensures $d\tau_N/d\tau_S < 0$ along S_1S_1.

But these costs are diminished by the lengthening of a period during which Southern firms earn positive profits of $\tilde{\pi}_S$ per consumer in the Northern market. Meanwhile, the marginal benefit to the South of prolonging its patents is greater than before, for somewhat the same reason. Since an increase in τ_S extends a phase in which Southern monopolists earn $M_S\pi_S + M_N\tilde{\pi}_S$, rather than just $M_S\pi_S$, it has a greater impact on the rate of Southern innovation than when $\tau_N > \tau_S$. The first-order condition that applies in place of (9) is

$$\left(C_c^{SS} - C_m^S - \pi_S - \frac{M_N}{M_S}\tilde{\pi}_S\right) + \omega(C_c^{NS} - C_m^N)$$
$$= \gamma_S\tilde{\Omega}_S\left[C_m^S + C_c^{SS}(\frac{\bar{T} - T_S}{T_S})\right] + \omega\gamma_N\Omega_S\left[C_m^N + C_c^{NS}(\frac{\bar{T} - T_S}{T_S})\right], \qquad (10)$$

where

$$\tilde{\Omega}_S = \frac{M_S T_S \pi_S + M_N T_S \tilde{\pi}_S}{(M_S T_S + M_N T_N)\,\pi_S + M_N(T_S - T_N)\tilde{\pi}_S}$$

and thus $\tilde{\Omega}_S > \Omega_S$ for $\tilde{\pi}_S > 0$.

The curve S_2S_2 in Figure 1 represents equation (10). It too must be downward sloping when the countries have identical Cobb-Douglas research technologies.[12] Notice that S_2S_2 is shifted to the right relative to a continuation of the curve S_1S_1. This displacement reflects the extra marginal benefit and reduced marginal cost of extending the patent length τ_S in situations where $\tau_S > \tau_N$ as compared to when $\tau_S < \tau_N$.

Each point on S_1S_1 gives a τ_S that locally maximizes U_S for the given value of τ_N.[13] Similarly, the points on S_2S_2 are locally optimal responses. To identify the *best* response to a given τ_N, we need to compare the welfare levels achieved at the local optima, as well as recognize the constraints on patent length. Consider, for example, the point at the intersection of S_1S_1 with the 45° line, where $\tau_N = \tau_1$. It is not

[12] In the figure, we depict a case in which the S_2S_2 curve hits the horizontal axis at a value of τ_S less than $\bar{\tau}$. In this case, the constraint on the length of the Southern patent — which can be no longer than the economic life of a product — never is binding. If the constraint does bind for some values of τ_N, the S_2S_2 curve will be situated above the horizontal axis at $\tau_N = 0$.

[13] This presumes that the second-order condition is satisfied for points on the curve, as in fact it must be if the countries have identical Cobb-Douglas research production functions.

optimal for the South's government to choose τ_1 in response to $\tau_N = \tau_1$, because the marginal benefit to the South of increasing its patent length is discretely larger than the marginal cost for τ_S slightly larger than τ_1. The South's best response to τ_1 is given by the relevant point on S_2S_2. Now consider the point of intersection of S_2S_2 with the 45° line, where $\tau_N = \tau_2$. By a similar argument, it is not optimal for the South to choose τ_2 in response to $\tau_N = \tau_2$, because the the marginal cost of extending the patent length exceeds the marginal benefit of doing so for τ_S slightly less than τ_2. In this case, the best response is to be found on the S_1S_1 curve. The South's best response function is represented by a boldface curve in Figure 1 with a discontinuity at $\tau_N = \tilde{\tau}$. The discontinuity comes at the value of τ_N at which the two local maxima yield the same level of utility.

We examine now the best responses in the North. Our procedure is the same. First, we consider the first-order condition for a locally optimal τ_N given τ_S and assuming $\tau_N > \tau_S$. Then we consider the first-order condition for a local optimum when $\tau_N < \tau_S$. Finally, we find the global optimum.

If $\tau_N > \tau_S$, an extension of the patent length in the North prolongs the period during which this country suffers deadweight loss on goods invented in the North and prolongs the period during which it faces monopoly prices for goods invented in the South. Meanwhile, the longer patents enhance the discounted profits from new innovations, and so augment the incentive for R&D in each region. Analogous to (9), we have the following condition equating the per-consumer marginal cost and marginal benefit:

$$
\begin{aligned}
&\left(C_c^{NS} - C_m^N - \pi_N\right) + \frac{1}{\omega}(C_c^{SS} - C_m^S) \\
&\qquad = \gamma_N \Omega_N \left[C_m^N + C_c^{NS}(\frac{\bar{T} - T_N}{T_N})\right] + \gamma_S \Omega_N \frac{1}{\omega}\left[C_m^S + C_c^{SS}(\frac{\bar{T} - T_N}{T_N})\right]. \qquad (11)
\end{aligned}
$$

This curve is depicted by N_1N_1 in Figure 1.

Like S_1S_1, the N_1N_1 curve must slope downward when the countries share identical Cobb-Douglas technologies in the research sector. More generally, we cannot rule out the possibility that the curve is upward sloping for part or all of its length. If the countries do have identical Cobb-Douglas research technologies and the S_1S_1 and N_1N_1 curves happen to intersect in the region in which $\tau_N > \tau_S$, then the absolute value of the slope of the former curve must exceed the absolute value of the slope of the latter at the point of intersection, as is required for the "stability" of a Nash equilibrium.[14]

When $\tau_S > \tau_N$, the first-order condition for maximizing U_N becomes

$$\begin{aligned}
&\left(C_c^{NN} - C_m^N - \pi_N\right) + \frac{1}{\omega}(\tilde{C}_m^S - C_m^S) \\
&\qquad = \gamma_N \Omega_N \left[C_m^N + C_c^{NN}\left(\frac{T_S - T_N}{T_N}\right) + C_c^{NS}\left(\frac{\bar{T} - T_S}{T_N}\right)\right] \\
&\qquad\qquad + \gamma_S \tilde{\Omega}_N \frac{1}{\omega}\left[C_m^S + \tilde{C}_m^S\left(\frac{T_S - T_N}{T_N}\right) + C_c^{SS}\left(\frac{\bar{T} - T_S}{T_N}\right)\right] \;, \qquad (12)
\end{aligned}$$

where $\tilde{\Omega}_N \equiv 1 - \tilde{\Omega}_S$. The first term on the left-hand side of (12) reflects the extra deadweight loss per Northern consumer that results from extending the period during which these consumers are served by a Northern monopolist rather than by competitive *Northern* producers. The second term reflects the loss per Northern consumer of surplus due to a marginally longer period of unconstrained monopoly pricing by Southern inventors instead of their being forced (perhaps) to engage in limit pricing. The first-term on the right-hand side of (12) represents the gain to a typical Northern consumer from the extra innovation that occurs thanks to the extended duration of monopoly profits for Northern inventors. The final term is the gain to a Northern consumer from the extra innovation that results when Southern firms are able to earn π_S instead of $\tilde{\pi}_S$ for a slightly longer period of time. If competition from the high-cost Northern producers does not induce any change in the pricing behavior

[14] By stability, we mean the usual tâtonnement adjustment process whereby τ_N adjusts to any local opportunities for welfare gain in the North and τ_S adjusts to such opportunities in the South. The claim about the relative slopes follows from the fact that, when the countries have identical Cobb-Douglas research technologies, γ_N and γ_S are identical constants and ω is independent of τ_S and τ_N. Then it is easy to see from (9) and (11) that the absolute value of the slope of S_1S_1 must be greater than $e^{\rho(\tau_N - \tau_S)} M_S/M_N$, while the absolute value of the slope of N_1N_1 must be less than $e^{\rho(\tau_N - \tau_S)} M_S/M_N$.

of Southern innovators (i.e., if $p_S \leq w_N a_S$), then the second terms on the left and right-hand sides of (12) both vanish.

Comparing (11) and (12), we see two differences. On the one hand, the marginal cost of extending the patent length is smaller (all else equal) when $\tau_S > \tau_N$ than when $\tau_N > \tau_S$, because $C_c^{NN} < C_c^{NS}$ and $\tilde{C}_m^S < C_c^{SS}$. By extending a patent when $\tau_N < \tau_S$, the North forestalls a gain in consumer surplus that is somewhat limited by the high cost of production in the North or by the continued proprietary position of a Southern inventor among potential producers in the South. On the other hand, the marginal benefit to the North of extending its patent also is smaller when $\tau_S > \tau_N$, because the lengthening of the patent protection matters somewhat less to a Southern innovator, who can earn the profit $\tilde{\pi}_S$ from each consumer in the North even without the benefit of a live patent there. Either the drop in the marginal cost or the drop in the marginal benefit might be larger at $\tau_N = \tau_S$. If the former is true, then the graph of (12) will look like that depicted in Figure 1; i.e., the intersection of the N_2N_2 curve with the 45° line will lie above and to the right of the intersection of the N_1N_1 curve with that line. If the drop in marginal benefit is bigger, the graph will look like the one shown in Figure 2. There, the intersection of the N_2N_2 curve with the 45° line lies below and to the left of the intersection of the N_1N_1 curve with that line.

In each figure, the North's best response function is depicted by a bold-faced curve. In Figure 1, the best response curve has three sections; it includes the entirety of N_1N_1, a segment of the 45° line, and the entirety of N_2N_2.[15] Consider, for example, the point where the N_1N_1 curve intersects the 45° line. If τ_N were approximately equal to τ_S but slightly larger than it, the first-order condition for local maximization of U_N would not quite be satisfied. The Northern government has a tiny incentive to reduce τ_N to the point where it equals τ_S. If it were to reduce τ_N a bit more, so that it became slightly smaller than τ_S, the marginal cost and marginal benefit of extending the patent length both would fall discretely; but, by the assumption

[15] If the intersection of the N_1N_1 curve with the vertical axis were to come at an ordinate greater than $\bar{\tau}$, then the North's best response function would have a fourth, horizontal segment along which the limit on the maximum patent length of $\bar{\tau}$ would be binding. Also, if the N_2N_2 curve were to hit the horizontal axis before $\tau_S = \bar{\tau}$, the best response function would include a segment with $\tau_N = 0$.

that underlies the figure, the former would fall by more. It would then pay for the Northern government to increase the duration of its patents back to the point where the countries' IPR protections are the same. For points along the bold portion of the 45° line, the Northern government has a strict incentive to reduce τ_N to equality with τ_S from all points above the line. But the Northern government does not have an incentive to cut τ_N to a length shorter than τ_S, because the sharp drop in marginal cost at the point where τ_S becomes larger than τ_N is enough to make the government want to increase τ_N back to equality with τ_S at points just below the 45° line. Finally, at the intersection of N_2N_2 with the 45° line, the first-order condition for local maximization of U_N is satisfied as the Northern government increases τ_N to equal τ_S from below. At points above the line, the marginal cost of an increase in τ_N exceeds the marginal benefit, so the government has incentive to reduce the patent duration back to the point of equality.

In Figure 2, the best response function for the North has two sections — a segment of the N_1N_1 curve and a segment of the N_2N_2 curve — and a discontinuity. Here the situation facing the Northern government is analogous to that facing the Southern government. At the intersection of N_1N_1 with the 45° line, the Northern government has a strict incentive to lower τ_N, because the marginal cost of extending the patent length exceeds by a discrete amount the marginal benefit of doing so for τ_N slightly smaller than this value of τ_S. At the intersection of N_2N_2 with the 45° line, the Northern government has an incentive to raise τ_N, because the marginal benefit of increasing the patent length rises sharply as we cross into region with $\tau_N > \tau_S$. The discontinuity comes at a value of τ_S where the two local maxima — on the N_1N_1 curve and on the N_2N_2 curve — yield the same level of aggregate welfare to residents of the North.

In each figure we depict a situation in which there exists a single Nash equilibrium characterized by longer patents in the North than in the South. It is not difficult to see that the situations illustrated in the figures are not the only possibilities. In each case, if the South's best response function were located somewhat further to the right, there might be two Nash equilibria, one with $\tau_N > \tau_S$ and another with $\tau_S > \tau_N$. A

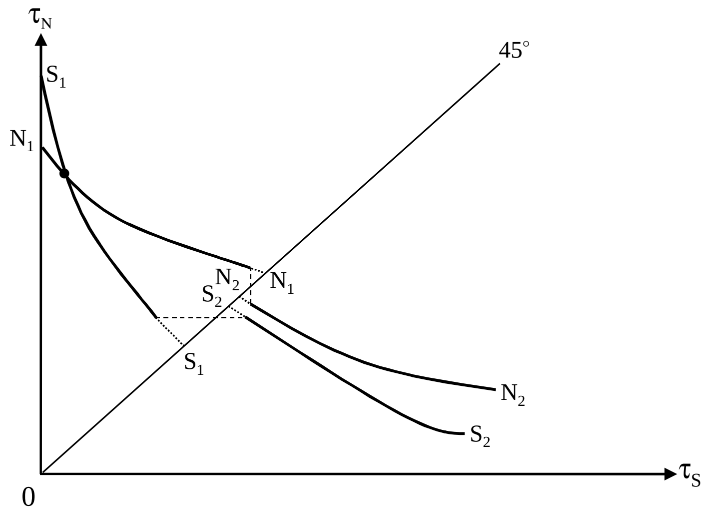

Figure 2: Nash Equilibrium with Discontinous Best Response in North

location of the South's best response function even further to the right might imply the existence of unique Nash equilibrium with $\tau_S > \tau_N$ or even $\tau_S = \bar{\tau}$. Moreover, for general production functions in the research sector, the various segments of the best response functions need not be everywhere downward sloping, nor must S_1S_1 for example be steeper than N_1N_1 at every point of intersection. Thus, in principle, there could be any number of Nash equilibria, some stable and others not. In a situation similar to that depicted in Figure 1, it seems there is still another possibility. If the S_1S_1 curve were to lie everywhere above the N_1N_1 curve for $\tau_N > \tau_S$ and the S_2S_2 curve were to lie everywhere below the N_2N_2 curve for $\tau_N < \tau_S$, then there would not exist any pure-strategy Nash equilibrium in the patent policy game.

There is not much that can be said in complete generality about the Nash equilibria of the policy game. One notable feature of any such equilibrium is that it does not involve the harmonization of patent policies. This observation follows immediately from the fact that the best response function of the low-wage country never touches the 45° line. For any patent policy that the North might choose, the South will either prefer that its patent lengths be strictly shorter than those in the North or that its patents be strictly longer than those in the North. This is because the South perceives a strictly higher marginal benefit from extending the duration of its patent when τ_S is slightly above τ_N than it does when τ_S is slightly below τ_N. When $\tau_S < \tau_N$, an increase in τ_S extends the period of positive profits for Southern producers only in the Southern market. But when $\tau_S > \tau_N$, an increase in τ_S extends a period of positive profits for these producers in both markets. Since the adverse effect on Southern consumers is the same in either case (if $\tau_S \approx \tau_N$), the Southern government never finds it optimal to mimic the patent policy of the North.

In reality, governments in the North typically grant longer patents and provide greater protection of intellectual property more generally than their counterparts in the South. In the next section, we will investigate how relative market size, relative endowments of human capital, and comparative advantages in manufacturing affect the pure-strategy Nash equilibria of the policy game, when such equilibria exist. Our goal is to understand why the North may have a greater incentive to grant long patents than the South, and to identify conditions under which there is a unique equilibrium with $\tau_N > \tau_S$.

4 Why are Patents Longer in the North?

In this section, we explore the determinants of the national patent policies in a non-cooperative Nash equilibrium. We do so by examining the comparative statics of our model, focusing especially on the case in which $F(H_j, L_{Rj}) = H_j^{1-\beta} L_{Rj}^{\beta}$ in country j, for $j = N, S$.

We begin with the sizes of the two markets. If M_S and M_N both grow at equal percentage rates, then there is no change in Ω_S, Ω_N, $\tilde{\Omega}_S$, or $\tilde{\Omega}_N$. By (9) and (10) there is then no effect on the S_1S_1 curve or the S_2S_2 curve. Similarly, by (11) and (12), the N_1N_1 and N_2N_2 curves remain in place. Also, there is no effect on the governments' preferred choices among the various local optima. It follows that an equi-proportionate expansion of the two markets leaves the Nash equilibrium patent policies unchanged. This extends our earlier finding for the closed economy: Balanced growth in the world market enhances the marginal cost and the marginal benefit of extending patents to similar extents and leaves the incentives facing the national governments in choosing their patent policies unaffected.

But changes in the *relative* size of the two markets do affect the national incentives for patent protection. Consider an increase in M_N/M_S. This decreases Ω_S and $\tilde{\Omega}_S$, while increasing Ω_N and $\tilde{\Omega}_N$. If the countries have Cobb-Douglas research technologies with similar exponents, then ω, γ_S and γ_N will be unaffected by a change in relative market size. In this case, the S_1S_1 curves shifts down and to the left, while the N_1N_1 and N_2N_2 curves shift up and to the right. As for the S_2S_2 curve, it may shift in either direction. For example, if $\tilde{\pi}_S$ is sufficiently small (but positive) or ω is sufficiently large, the curve will shift to the left with increases in the relative size of the North if M_N/M_S initially is small, but will shift to the right with further increases in the relative size of the North if M_N/M_S already is large.

We organize our discussion of the implications for the equilibrium policies around the following proposition:

Proposition 1 *Let the countries share identical Cobb-Douglas research technologies.*
(i) Suppose that for $M_N/M_S = m'$ there exists a unique Nash equilibrium with $\tau'_N > \tau'_S$ and for $M_N/M_S = m'' > m'$ there exists a unique Nash equilibrium with $\tau''_N > \tau''_S$. Then $\tau''_N > \tau'_N$ and $\tau''_S < \tau'_S$.
(ii) In the limit, as $M_N/M_S \to \infty$, there is a unique Nash equilibrium with $\tau_S = \bar{\tau}$.

The first part of the proposition says that an increase in the relative size of the Northern market increases equilibrium patent protection in the North and decreases it in the South, provided that the equilibrium is unique before and after the change and that Northern patents are longer in each case. This result follows directly from the fact that the N_1N_1 curve shifts up and the S_1S_1 curve shifts to the left, as M_N/M_S grows. The second part of the proposition says that, when the Southern market is negligibly small compared to that in the North, the Southern government always sets a maximal patent length equal to the economic life of a differentiated product. This follows from the fact that the marginal cost of extending the patent becomes negative and large as M_N/M_S approaches infinity. We will say more about this in a moment.

Why does a change in the relative size of the markets change the relative incentives that the governments face in choosing their patent policies? The answer has to do with the effects of such a change on the marginal cost and marginal benefit of extending the length of patents in each country. Suppose, for example, that there is a unique equilibrium with $\tau_N > \tau_S$ and that M_N rises. This has no effect on the marginal cost of longer patent protection in the South (provided that we remain in a regime with $\tau_N > \tau_S$). But since monopolists in both countries now realize a smaller share of their total profits from sales in the South, the responsiveness of innovation in each country to changes in τ_S falls.[16] It follows that the marginal benefit of extending the Southern patent falls, and the South has less incentive to grant lengthy patents, for any given patent duration in the North. In the North, the expansion of market size causes a directly proportional increase in the marginal cost and marginal benefit of longer patents, since both the deadweight loss from monopoly and the consumer gains from greater innovation are realized by more households. But the marginal benefit gets an added boost, because innovation in both countries becomes more responsive to the Northern patent policy.[17] Thus, the Northern government has reason to lengthen its patents, for any given patent policy in the South.

However, for M_N/M_S large enough, it becomes a dominant strategy for the government of the South to choose a patent length equal to the full economic life of a product. The explanation for this is clear. For large M_N/M_S, the profit opportu-

[16] Take, for example, the case in which $\tau_N > \tau_S$. In this case,

$$\frac{1}{v_j}\frac{dv_j}{d\tau_S} = \frac{\rho e^{-\rho\tau_S}}{(1-e^{-\rho\tau_S}) + \frac{M_N}{M_S}(1-e^{-\rho\tau_N})}.$$

Then an increase in M_N/M_S reduces both $(1/v_N)(dv_N/d\tau_S)$ and $(1/v_S)(dv_S/d\tau_S)$. But these semi-elasticities determine the elasticity of research with respect to patent length in each country.

[17] For example, with $\tau_N > \tau_S$,

$$\frac{1}{v_j}\frac{dv_j}{d\tau_N} = \frac{\frac{M_N}{M_S}\rho e^{-\rho\tau_S}}{(1-e^{-\rho\tau_S}) + \frac{M_N}{M_S}(1-e^{-\rho\tau_N})},$$

which is an increasing function of M_N/M_S. It follows that $(1/\dot{n}_N)(d\dot{n}_N/d\tau_N)$ and $(1/\dot{n}_S)(d\dot{n}_S/d\tau_N)$ also are increasing functions of M_N/M_S.

nities for Southern firms in the North dwarf considerations of consumer surplus in the South. The welfare of the South then is maximized by whatever policy gives Southern inventors the greatest discounted profits in the Northern market. But the profit-maximizing patent policy is one that grants proprietary rights to an innovator for the entire economic life of his product. A patent length shorter than $\bar{\tau}$ means that a Southern inventor will face competition at the expiration of the patent from compatriot producers. But such competition would spell a terms of trade loss for the South with only a negligible gain in consumer surplus to compensate. Thus, a relatively small market in the South tends to favor short-lived patents there, but only up to a point!

Next we consider the relative endowments of human capital, our proxy for the relative capacities for research in the two countries. We assume for this purpose that the demand for innovative products has a constant elasticity, as is the case when $h(x) = x^{\alpha}$ for $\alpha \in (0, 1)$. First note that with Cobb-Douglas research technologies (and many others), an increase in H_N/H_S must increase ω, the relative research output of the North.[18] So we investigate the comparative statics of the system of (9) and (11) and the system of (10) and (12) with respect to changes in ω.

An increase in ω shifts the S_1S_1 curve to the left, because the marginal cost of extending patents in the South rises relative to the marginal benefit from doing so.[19] Intuitively, the marginal cost of increasing τ_S is sensitive to changes in ω, because there is no profit offset to the loss of Southern surplus on Northern inventions as there is for Southern inventions. Meanwhile, an increase in ω shifts the N_1N_1 curve

[18] With $\phi_j = H_j^{1-\beta} L_{Rj}^{\beta}$ in country j, for $j = N, S$,

$$\omega = \frac{H_N}{H_S}\left(\frac{b_N v_N}{b_S v_S}\right)^{\frac{\beta}{1-\beta}} .$$

Since v_N/v_S is independent of H_N/H_S for $\tau_N > \tau_S$, ω is an increasing, linear function of H_N/H_S.

[19] With constant elasticity demands, $C_c^{NS}/C_m^N > C_c^{SS}/C_m^S$. This is enough to ensure that

$$\frac{C_c^{NS} - C_m^N}{C_c^{SS} - C_m^S - \pi_S} > \frac{C_m^N T_S + C_c^{NS}(\bar{T} - T_S)}{C_m^S T_S + C_c^{SS}(\bar{T} - T_S)} ,$$

which in turn implies that the left-hand side of (9) increases with ω by more than the right-hand side.

upward if w_N is close to w_S, but downward if w_N is much larger than w_S. In the former case, the marginal cost of extending patents falls relative to the marginal benefit, because the profits earned by Northern producers are an offset to the surplus loss from patents when goods originate in the North. In the latter case, the Northern profits are small, and the difference between marginal benefit and marginal cost falls with ω, because Northern inventions generate a smaller discounted surplus for consumers over their economic lives. It follows that if the initial equilibrium has $\tau_N > \tau_S$ and w_N close to w_S, then an increase in the relative skill endowment in the North must lengthen patents in the North and shorten those in the South. Whereas, if the initial equilibrium has $\tau_N > \tau_S$ and w_N much larger than w_S, then an increase in the relative skill endowment of the North may induce an increase in τ_S or a fall in τ_N. Still, if $M_N \geq M_S$, an intersection of the S_1S_1 curve and the N_1N_1 in the region with $\tau_N > \tau_S$ is guaranteed for H_N/H_S (and thus ω) large enough.

For cases in which the initial equilibrium has $\tau_S > \tau_N$, we must look at the shifts in S_2S_2 and N_2N_2 that are induced by an increase in ω. We show in the appendix that with identical Cobb-Douglas research technologies and constant elasticity demands, the S_2S_2 curve must shift to the left and the N_2N_2 curve must shift upward (or remain in place) when ω increases. It follows that τ_S must fall and τ_N must rise in response to a rise in H_N/H_S when the initial equilibrium has longer patents in the South.

This suggests that an equilibrium with longer patents in the South may not exist at all if the relative research capacity of the North is sufficiently large. In fact, we can prove

Proposition 2 *If $M_N \geq M_S > 0$ and H_N/H_S is sufficiently large, then $\tau_N > \tau_S$ in any Nash equilibrium.*

The proof involves our showing that (10) and (12) cannot simultaneously hold when $M_N \geq M_S$, $T_S \geq T_N$ and ω approaches infinity. As $\omega \to \infty$, the dominant term in the South's marginal cost becomes $C_c^{NS} - C_m^N$. The dominant term in the North's marginal cost becomes $C_c^{NN} - C_m^N - \pi_N$, which is smaller, since $C_c^{NN} < C_c^{NS}$ and $\pi_N > 0$. Meanwhile, the dominant term in the South's marginal benefit becomes

$\gamma_N \Omega_S [C_m^N + C_c^{NS}(\bar{T} - T_S)/T_S]$, while that in the North's marginal benefit becomes $\gamma_N \Omega_N [C_m^N + C_c^{NN}(T_S - T_N)/T_N + C_c^{NS}(\bar{T} - T_S)/T_N]$. It is straightforward to check that when $M_N \geq M_S$, the latter term must be the larger of the two when $T_S \geq T_N$. Thus, the two first-order conditions cannot both be satisfied for any $\tau_S \geq \tau_N$.

We have thus identified conditions under which an equilibrium must be characterized by longer patents in the North than in the South. These conditions describe a market for innovative products that is at least as large in the North as in the South, and an inventive capacity in the North that is much greater than that in the South. More generally, the greater is the research capacity in the North relative to that in the South, the more likely it is that there will be a unique equilibrium with $\tau_N > \tau_S$. An increase in H_N/H_S increases aggregate profit income in the North relative to that in the South. Since the monopoly profits are an offset to the consumer cost of patent protection, the North's relative incentive to provide proprietary rights grows as its firms' relative share of world profits increases.

Finally, we consider the countries' comparative advantages. Recall that the production of the homogeneous good requires b_j units of labor per unit of output in country j, while differentiated products invented in country j require a_j units of labor per unit of output. We use $\kappa_j \equiv b_j (a_j)^{-\alpha}$ as a measure of country j's comparative advantage in differentiated products, where α is the exponent in the consumers' utility function; i.e., $h(x) = x^\alpha$. With Cobb-Douglas research technologies and constant elasticity demands, an increase in κ_N/κ_S induces an increase in ω, the relative research output of the North.[20] But we have already seen that an increase in ω reduces the likelihood that there exists an equilibrium with $\tau_S \geq \tau_N$. Thus, comparative advantage in differentiated products is yet another factor that works in favor of longer patents in the North.

[20]See the expression for ω in footnote 18. A larger value of b_N/b_S implies a smaller relative wage in the North. This reduces the relative cost of R&D there, thus raising its relative profitability. A lower value of a_N/a_S spells higher relative profits for a Northern monopoly, and thus a relatively greater value for Northern inventions. We find that v_j is proportional to $(b_j/a_j)^{\alpha/(1-\alpha)}$, and thus ω is an increasing function of κ_N/κ_S.

5 International Patent Agreements

In this section, we will study international patent agreements. We begin by characterizing the combinations of patent policies that are jointly efficient for the two countries.[21] For the most part, we shall assume that the countries can offer and receive direct compensation — corresponding either to financial payments or to welfare transfers that are effected by concessions in policy areas other than IPR. Then the efficient patent regimes are those combinations of τ_S and τ_N that maximize the sum of the countries' welfare levels gross of any transfers. Arguably, an international negotiation can be used to achieve some such outcome, if the countries are discussing several issues simultaneously as part of a comprehensive trade negotiation and if there are no frictions in the bargaining process.

After we have identified the efficient policy combinations, we examine the inefficiencies that are present in a noncooperative equilibrium with $\tau_N > \tau_S$. By comparing the Nash equilibrium outcomes with the efficient policies, we can point to the changes in the patent regime that ought to be effected by an international treaty.

The final part of this section deals with the question of policy harmonization. By that point, we will have shown that harmonization is not necessary for global efficiency. We will 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.

5.1 Efficient Patent Regimes

We begin with the assumption that countries can effect international transfers or provide compensation in other policy areas, so that the efficient patent lengths are ones that maximize the joint welfare of the two countries. We focus here on regimes with $\tau_N \geq \tau_S$. A regime with $\tau_S > \tau_N$ typically will not be efficient in our setting, because

[21] 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.

such a regime involves one and probably two extra distortions. First, when $\tau_S > \tau_N$, there is a period between the expiration of the two patents when a good invented in the North is produced competitively in that country. Since the original inventor earns no profits during this period of local production, there is no encouragement of R&D relative to the situation that would exist if these goods instead were produced in the South. But the Northern consumers enjoy less surplus with production in the North than they would if production were done in the South, which reveals an excess burden of the patent regime with $\tau_S > \tau_N$. Second, a regime with $\tau_S > \tau_N$ has relatively more R&D taking place in the South than does a regime with $\tau_N \geq \tau_S$. But, as we will see below, in an efficient regime with $\tau_N \geq \tau_S$ the social marginal product of labor in the North's research sector typically exceeds that in the South's research sector. If it does, it would be costly to have an international patent regime with $\tau_S > \tau_N$ that gives still greater relative incentive for innovation in the South.

Consider the efficient choice of patent policies τ_N and τ_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). If $\tau_N \geq \tau_S$, aggregate welfare in the South is given by

$$U_S(0) = \Lambda_{S0} + \frac{w_S(L^S - L_R^S)}{\rho} + \frac{M_S\phi_S}{\rho}\left[T_S C_m^S + (\bar{T} - T_S)C_c^{SS}\right]$$
$$+\frac{M_S\phi_N}{\rho}\left[T_S C_m^N + (\bar{T} - T_S)C_c^{NS}\right] + \frac{\phi_S}{\rho}\pi_S\left(M_S T_S + M_N T_N\right) , \qquad (13)$$

where Λ_{S0} is the fixed amount of discounted surplus that consumers in the South derive from goods that were invented before time 0. For the North,

$$U_N(0) = \Lambda_{N0} + \frac{w_N(L^N - L_R^N)}{\rho} + \frac{M_N\phi_S}{\rho}\left[T_N C_m^S + (\bar{T} - T_N)C_c^{SS}\right]$$
$$+\frac{M_N\phi_N}{\rho}\left[T_N C_m^N + (\bar{T} - T_N)C_c^{NS}\right] + \frac{\phi_N}{\rho}\pi_N\left(M_S T_S + M_N T_N\right) , \qquad (14)$$

where Λ_{N0} is defined analogously to Λ_{S0}. In both (13) and (14), the second term is the income from home production of homogeneous goods, the third and fourth terms are the discounted surplus generated by goods invented in the South and the North, respectively, and the last term is the country's discounted sum of profit income.

Summing (13) and (14), we find that

$$\rho\left[U_S(0)+U_N(0)\right]=\rho\left(\Lambda_{S0}+\Lambda_{N0}\right)$$
$$+w_S(L^S-L_R^S)+w_N(L^N-L_R^N)+(M_S+M_N)\,\bar{T}\left(\phi_S C_c^{SS}+\phi_N C_c^{NS}\right)$$
$$-\left(M_S T_S+M_N T_N\right)\left[\phi_S\left(C_c^{SS}-C_m^S-\pi_S\right)+\phi_N\left(C_c^{NS}-C_m^N-\pi_N\right)\right] \quad (15)$$

Also, $v_j = \pi_j(M_S T_S + M_N T_N)$. It follows that changes in τ_S and τ_N that leave $M_S T_S + M_N T_N$ unaffected leave v_S and v_N unaffected, and therefore also leave L_R^S, L_R^N, ϕ_S and ϕ_N unaffected. Such changes have no effect on the allocation of labor in either country and no effect on aggregate world welfare.[22]

The variable $Q = M_S T_S + M_N T_N$ measures the total protection afforded to creators of intellectual property by the international patent system. It 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. The same global reward for innovation can be achieved with different combinations of the two patent lengths. Evidently, the combinations also generate the same aggregate levels of world income and consumer surplus. However, the distribution of world surplus is not invariant to the specific national policies. The longer are patents in the North, the shorter and more delayed is the period during which the Northern consumers enjoy the benefits of competitive pricing for a good. Similarly, the shorter are patents in the South, the sooner and longer will consumers enjoy a high level of surplus there. For a given value of Q, the South fares better and the North worse (absent any compensating transfers) the greater is τ_N and the smaller is τ_S.

A globally efficient patent regime has $M_S T_S + M_N T_N = Q^*$, where Q^* is the value of Q that maximizes the right-hand side of (15).[23] Notice that a range of

[22]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.

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

$$C_c^{SS}-C_m^S-\pi_S+\omega\left(C_c^{NS}-C_m^N-\pi_N\right)$$
$$=\left[\frac{(M_S+M_N)\bar{T}-Q^*}{Q^*}\right]\left(\gamma_S C_c^{SS}+\omega\gamma_N C_c^{NS}\right)+\gamma_S C_m^S+\omega\gamma_N C_m^N\ .$$

efficient outcomes can be achieved without the need for international transfers. By appropriate choice of τ_N and τ_S with $\tau_N \geq \tau_S$ and $M_S T_S + M_N T_N = Q^*$, 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 $T_S = T_N = Q^*/(M_S + M_N)$. For distributions of world welfare in this range, the efficient patent regime is the same whether international transfers are feasible or not. Only if it were desirable to give still more welfare to the North than this country enjoys when the patent lengths are such that $T_S = T_N = Q^*/(M_S + M_N)$ would the feasibility of international transfers make any difference to the efficient allocations.

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 3. The figure depicts the best response functions and the efficient policy combinations in the space of T_S and T_N. The efficient combinations are those for which $M_S T_S + M_N T_N = Q^*$, as represented by the points on the line labelled QQ. The South's best response function for $T_N \geq T_S$ is given by (9). In the Cobb-Douglas case, this too is a line, as represented by SS in the figure. Similarly, the North's best response function is described by (11), which is represented by NN in the figure.

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

We can also calculate from (15) the social marginal product of labor in research in each country compared to its social marginal product in producing homogeneous goods. We find this ratio to be $(C_c^{jS} - C_m^j - \pi_j)/\pi_j$ in country j. For constant elasticity and many other demand functions, this ratio is higher for the North than it is for the South. This validates our claim that it would often be beneficial to world welfare if the composition of world research could be changed in favor of relatively more innovation in the North.

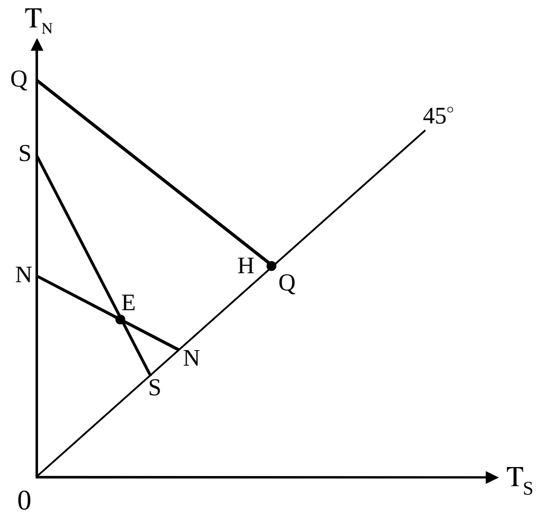

Figure 3: Efficient Harmonization versus Nash Equilibrium

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 extra monopoly profits to Northern innovators, which contributes to aggregate income there. Second, an increase in τ_S enhances the incentives for R&D, inducing an increase in both ϕ_S and ϕ_N. The extra product diversity that results from this R&D creates additional surplus for Northern consumers.

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 curve must lie outside any Nash equilibrium with $T_N \geq T_S$. We record our finding in

Proposition 3 *Let (τ_S, τ_N) be any Nash equilibrium in the policy game with $\tau_N \geq \tau_S$ and let (τ_S^*, τ_N^*) be an efficient combination of patent policies. Then*

$$M_S(1 - e^{-\rho\tau_S^*}) + M_N(1 - e^{-\rho\tau_N^*}) > M_S(1 - e^{-\rho\tau_S}) + M_N(1 - e^{-\rho\tau_N}).$$

The proposition implies that, starting from a Nash equilibrium in which patents are longer in the North than in the South, an efficient 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 often 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 efficient 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 τ_S is a best response function for 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 policy changes, we focus on a treaty that would achieve global effi-

[24] See also Lai and Qiu (2000), 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.

ciency. Such a treaty is represented by point H in Figure 3. 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), but in general it might entail either an increase or a decrease of patent duration in the North. In the case of identical Cobb-Douglas research technologies, however, point H lies above the intersection of the NN curve with the vertical axis, as drawn. 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 $T_S = 0$. Since the NN curve is downward sloping in the Cobb-Douglas case, the fact that it starts below point H implies that the efficient harmonized patents are longer in both countries than the patents chosen in a Nash equilibrium.

Among all policy combinations that achieve global efficiency, the harmonized policies are the ones that provide the greatest benefit to the North and the least benefit to the South. Moreover, 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. The South certainly would be harmed by efficient harmonization if the research technologies were other than Cobb-Douglas and the initial Nash equilibrium happened to be at a point above and to the left of H. Then both the increase in the length of Southern patents and the decrease in the length of Northern patents would work to the detriment of the South. In this case, the North definitely gains from efficient harmonization.[25] In general, the larger are M_N/M_S and ω, the more likely it is that the North would gain and the South would lose from efficient harmonization. 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.[26]

[25] It is theoretically possible that the South would gain and the North would lose from a patent agreement calling for efficient harmonization. However, it can be shown that if $M_N \geq M_S$ and $\omega \geq \pi_S/\pi_N$, the North is sure to benefit from a move from the non-cooperative equilibrium to the point of efficient harmonization.

[26] McCalman (2000) estimate the income transfers implicit in the TRIPs agreements and finds that international patent harmonization benefits the United States at the expense of the developing

6 Conclusions

We have developed a simple model of endogenous innovation and have used it to study the incentives facing national governments in choosing their patent policies. Our 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 relating 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, countries face different incentives in setting their national policies due to differences in factor prices, market sizes, and capacities for doing research. 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 policy is found by equating the sum of extra deadweight loss that results from extending patents for domestic firms and the extra surplus loss that results from extending monopoly pricing by foreign firms with the benefits that flow from providing greater incentives for innovation for firms in both countries. A country's optimal patent length depends on the policies chosen by its trading partner, because the foreign policy affects the global incentives for innovation and the relative numbers of patent-holders in each country.

Our analysis revealed a subtlety that arises whenever factor prices differ across countries. When patents are longer in the high-wage North, a lengthening of patents in the South postpones a period of competitive pricing by low cost producers. But when patents are longer in the low-wage South, a lengthening of patents in the North postpones a period of limit pricing by Southern innovators and a period of competitive pricing of goods invented in the North by high cost producers in that country. Due to this asymmetry between a patent regime with longer patents in the North and one with longer patents in the South, it is possible to have multiple equilibria in the policy game, or for there not to exist any pure-strategy equilibrium at all.

countries and of Canada, the United Kingdom and Japan.

We find 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 protection typically increases as its relative capacity for research grows. In a noncooperative equilibrium, patents must be longer in a country whose relative capacity for research is sufficiently great, provided that its market size is no smaller than that of its trade partner. Thus, the large markets in the North and the relatively greater capacity for research there can explain why these countries often have stronger protection of intellectual property than their Southern trading partners.

Starting from a Nash equilibrium, the two governments will have an incentive to negotiate an international patent agreement. An agreement can ensure that national policies will 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 call for a lengthening of patents in at least one country. The 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 the non-cooperative equilibrium typically would benefit the North quite possibly at the expense of the South.

Our analysis can be extended to more general environments. Two extensions come readily to mind. First, we have assumed that firms in the North must produce their innovative products locally, despite the lower wages that prevail in the South. We could easily modify the model to allow for direct foreign investment, and examine how such investment affects the national incentives for patent protection. Direct foreign investment (DFI) by Northern firms — when it occurs — increases the profits π_N that flow to a Northern inventor, and (in most cases) the surplus C_m^N that accrues while a Northern firm holds a live patent. When $\tau_N > \tau_S$, the effect is to raise the marginal benefit of extending patents and reduce the marginal cost of doing so in both countries.

Second, we have assumed that the countries have similar demands for innovative products, except perhaps for differences in market size. If demand functions were to differ in the two countries, the deadweight loss that results from patent protection also would differ. Differences in demand would also be reflected in the characteristics of globally efficient patent regime. An efficient patent regime would equalize the marginal deadweight loss in the two countries associated with providing a given push to global innovation.

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

[4] Lai, Edwin L.-C. and Qiu, Larry D. (2000), The North's Intellectual Property Rights Standard for the South?" City University of Hong Kong, manuscript.

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

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

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

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

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

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

[11] 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.

7 Appendix

In this appendix we examine how changes in ω shift the S_2S_2 and N_2N_2 curves. We focus on the case in which the North and the South share identical Cobb-Douglas research technologies and the demands for the differentiated products have constant elasticities.

7.1 Shifts in S_2S_2

We will prove that the S_2S_2 curve shifts to the left when ω increases. To this end, note that the right-hand side of (10) is a decreasing function of τ_S. We will show that, starting from a point on the S_2S_2 curve, where the left-hand and right-hand sides of (10) are of course equal, an increase in ω increases the left-hand side by more than it increases the right-hand side. Then a reduction in τ_S will be needed (at given τ_N) to restore equality between the two sides of the equation.

Let $R_S(\omega;\tau_S,\tau_N)$ be the ratio of the right-hand side of (10) to the left-hand side of (10). With $\gamma_S=\gamma_N$, a necessary and sufficient condition for $\partial R_S/\partial\omega<0$ is

$$\frac{\tilde{\Omega}_S}{\Omega_S}\left[\frac{C_m^S T_S + C_c^{SS}(\bar{T}-T_S)}{C_c^{SS}-C_m^S-\pi_S-\frac{M_N}{M_S}\tilde{\pi}_S}\right] > \frac{C_m^N T_S + C_c^{NS}(\bar{T}-T_S)}{C_c^{NS}-C_m^N}\ . \tag{A1}$$

But $\tilde{\Omega}_S>\Omega_S$ and $\pi_S+(M_N/M_S)\tilde{\pi}_S>0$, so (A1) will be satisfied if

$$\frac{C_m^S T_S + C_c^{SS}(\bar{T}-T_S)}{C_c^{SS}-C_m^S} \geq \frac{C_m^N T_S + C_c^{NS}(\bar{T}-T_S)}{C_c^{NS}-C_m^N}$$

or

$$\left(C_c^{NS}-C_m^N\right)\left[C_m^S T_S + C_c^{SS}(\bar{T}-T_S)\right] - \left(C_c^{SS}-C_m^S\right)\left[C_m^N T_S + C_c^{NS}(\bar{T}-T_S)\right] \geq 0\ . \tag{A2}$$

The left-hand side of (A2) can be written as $\bar{T}(C_m^S C_c^{NS}-C_m^N C_c^{SS})$. With constant elasticity demands, this must be positive, because $C_m^S/C_c^{SS}=C_m^N/C_c^{NN}>C_m^N/C_c^{NS}$. It follows that an increase in ω shifts the S_2S_2 curve to the left.

7.2 Shifts in N_2N_2

To see how an increase in the relative innovation rate in the North affects the N_2N_2 curve, we focus on the case of a moderate gap between the wages in the North and in the South. In such circumstances, a Southern innovator will be forced to reduce its price in the North at the time that its patent expires there in order to deter potential competition from Northern imitators.[27] With a constant elasticity of demand for differentiated products, $h(x) = x^\alpha$ for $\alpha \in (0,1)$. Then a price cut occurs after the expiration of a patent in the North whenever $w_N/w_S < 1/\alpha$.

We aim to show that the N_2N_2 curve shifts up with increases in ω. At an initial value of ω, the left and right-hand sides of (12) are equal for any point on N_2N_2. We need to establish that, starting from such a point, an increase in ω decreases the left-hand side of (12) by more than it decreases the right-hand side. Then an increase of τ_N — which reduces the right-hand side without changing the left-hand side — will be needed to restore the equality between perceived marginal cost and marginal benefit of lengthening patents in the North.

Let $R_N(\omega;\tau_S,\tau_N)$ denote the ratio of the right-hand side of (12) to the left-hand side of (12). With $\gamma_S = \gamma_N$, a necessary and sufficient condition for $\partial R_N/\partial\omega > 0$ is

$$\frac{\tilde{\Omega}_N}{\Omega_N}\left[\frac{C_m^S T_N + \tilde{C}_m^S(T_S - T_N) + C_c^{SS}(\bar{T} - T_S)}{C_m^N T_N + C_c^{NN}(T_S - T_N) + C_c^{NS}(\bar{T} - T_S)}\right] < \frac{\tilde{C}_m^S - C_m^S}{C_c^{NN} - C_m^N - \pi_N}$$

or

$$\frac{\tilde{\Omega}_N C_c^{SS}}{\Omega_N C_c^{NS}}\left[\frac{\bar{T} - \left(1 - \frac{\tilde{C}_m^S}{C_c^{SS}}\right)T_S - \left(\frac{\tilde{C}_m^S - C_m^S}{C_c^{SS}}\right)T_N}{\bar{T} - \left(1 - \frac{C_c^{NN}}{C_c^{NS}}\right)T_S - \left(\frac{C_c^{NN} - C_m^N}{C_c^{NS}}\right)T_N}\right] < \frac{\tilde{C}_m^S - C_m^S}{C_c^{NN} - C_m^N - \pi_N}\,. \tag{A3}$$

Note that $\tilde{C}_m^S = C_c^{SN}$. Also, with constant elasticity demand, $C_c^{SN}/C_c^{SS} = C_c^{NN}/C_c^{NS}$ and $C_m^S/C_c^{SS} = C_m^N/C_c^{NN}$. These equalities, plus $C_c^{NN} < C_c^{NS}$ imply that the left-hand side of (A3) is an increasing function of both T_N and T_S. Therefore, we can

[27] If the Southern innovators charge the same price in the North before and after the expiration of their patents there, then $C_m^S = \tilde{C}_m^S$ and $\pi_S = \tilde{\pi}_S$. But the latter equality implies that $\tilde{\Omega}_N = 0$. Therefore, ω does not appear in the equation for the N_2N_2 curve. Clearly, a change in ω has no effect on the North's choice of patent policy (for $\tau_N < \tau_S$) in this case.

replace T_N and T_S in this term by the larger number $\bar{T}$ to obtain a sufficient condition for $\partial R_N/\partial\omega > 0$. We can also replace $\tilde{\Omega}_N/\Omega_N$ by $(\pi_S - \tilde{\pi}_S)/\pi_S$, because $\tilde{\Omega}_N/\Omega_N < (\pi_S - \tilde{\pi}_S)/\pi_S$. After making these substitutions, the resulting sufficient condition is

$$\left(\frac{\pi_S - \tilde{\pi}_S}{\pi_S}\right)\frac{C_m^S}{C_m^N} < \frac{\tilde{C}_m^S - C_m^S}{C_c^{NN} - C_m^N - \pi_N}$$

or

$$\tilde{\pi}_S\left(C_m^S C_c^{NN} - C_m^N C_m^S\right) - \pi_S\left(C_m^S C_c^{NN} - C_m^N \tilde{C}_m^S\right) + (\pi_S - \tilde{\pi}_S)\,\pi_N C_m^S > 0\ . \tag{A4}$$

The third term on the left-hand side of (A4) clearly is positive, so we focus on the difference between the first term and the second term. This difference is positive if and only if

$$\frac{\tilde{\pi}_S}{\pi_S} > \frac{C_m^S C_c^{NN} - C_m^N C_m^S}{C_m^S C_c^{NN} - C_m^N \tilde{C}_m^S}\ .$$

With constant elasticity demand,

$$\frac{\tilde{\pi}_S}{\pi_S} = \frac{\alpha}{1-\alpha}\left(\frac{w_N}{w_S} - 1\right)\left(\frac{\alpha w_N}{w_S}\right)^{-\frac{1}{1-\alpha}}$$

and

$$\frac{C_m^S C_c^{NN} - C_m^N C_m^S}{C_m^S C_c^{NN} - C_m^N \tilde{C}_m^S} = \frac{1 - \left(\frac{w_N}{w_S}\right)^{-\frac{\alpha}{1-\alpha}}}{1 - \left(\frac{1}{\alpha}\right)^{-\frac{\alpha}{1-\alpha}}}\ .$$

It is straightforward to verify that

$$\frac{\alpha}{1-\alpha}\left(\frac{w_N}{w_S} - 1\right)\left(\frac{\alpha w_N}{w_S}\right)^{-\frac{1}{1-\alpha}} > \frac{1 - \left(\frac{w_N}{w_S}\right)^{-\frac{\alpha}{1-\alpha}}}{1 - \left(\frac{1}{\alpha}\right)^{-\frac{\alpha}{1-\alpha}}}$$

for all $w_N/w_S \in (1, 1/\alpha)$. Therefore, inequality (A4) is satisfied, which is sufficient to ensure that (A3) is satisfied. This establishes that an increase in ω shifts the N_2N_2 upward when $w_N/w_S < 1/\alpha$.