# INTELLECTUAL PROPERTY AND THE SCALE OF THE MARKET

MICHELE BOLDRIN AND DAVID K. LEVINE

*Departments of Economics University of Minnesota and UCLA*

ABSTRACT. Intellectual property protection involves a tradeoff between the undesirability of monopoly and the desirable encouragement of creation and innovation. As the scale of the market increases, due either to economic growth, or the expansion of intellectual property rights through treaties such as the World Trade Organization, this tradeoff changes. We show that generally speaking, the socially optimal amount of protection decreases as the scale of the market increases.

Keywords: Economic Theory, Intellectual Property, Industrial Organization, Monopoly.

JEL Classification: X15; X16

We thank the National Science Foundation for financial support. Our ideas benefited from comments at the Theory Workshop at Columbia University and from a discussion with Kyle Bagwell.

Corresponding Author: David K. Levine, Department of Economics, UCLA, Los Angeles, CA 90095, USA. Phone/Fax: 310-825-3810. Email: david@dklevine.com.

First Version: 2nd December, 2003, This Version: 17th May 2004.

## 1. Introduction

The U.S. Constitution gives Congress the power to "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." This recognizes the two basic economic features of intellectual property protection: on the one hand exclusive rights create monopoly power and so should be limited in time. On the other, monopoly power provides an incentive for creation and innovation. For practical reasons the same time limit applies across a wide variety of different creations and innovation: In U.S. law, copyright is life of author plus 70 years for individual works; 95 years for works for hire; design patents are 20 years and ornamentation patents are 14 years. Since the private profitability of creating and innovating varies widely, this means that for any fixed time limit many ideas will earn profits above and beyond the level needed to recoup the cost of innovation. In a larger market profits will be greater, and inframarginal ideas will earn additional economically unnecessary rents. Obviously, as the market expands, it becomes possible to reduce the length of term without reducing the production of new ideas. But, as the market expands, not only will some ideas earn economically unnecessary rents, but some ideas that were not profitable to produce will become so, and reducing the length of term will discourage these marginal entrepreneurs. In this paper we look at the general equilibrium interaction among these forces.

Our basic result is an intuitive one. Optimal copyright and patent term length involves a tradeoff between increasing the unnecessary consequences of monopoly on inframarginal ideas, and increasing the number of marginal ideas. As the scale of the market increases, it will generally be desirable to give up some of the additional marginal ideas in exchange for reduction of monopoly across the broad variety of ideas that will be produced anyway, and so the optimal policy should reduce the length of protection as the scale of the market increases.

We make this point in the context of a simple static general equilibrium model of innovation and creation, which is designed explicitely to allow intellectual property protection to have a beneficial social role. Ideas are created subject to an indivisibility. There are many possible ideas, and to model the fact that each should have downward sloping demand, we adopt the Dixit-Stiglitz model of preferences. Ideas differ in quality. We consider first the case in which quality has a neutral effect on the relationship between the private and social benefit of an idea, and the production of ideas requires only a fixed cost. When the market is sufficiently small, it may be optimal to provide an unlimited monopoly, but we show that when the market is large enough, a time limit should always be imposed, and this limit should strictly decrease as the size of the market grows. Although there may be some intermediate regions of market size where the socially optimal limit increases rather than decreases with size, this is does not happen in the most plausible case in which the elasticity of revenue increases with quality.

One point is important to emphasize. Increasing the scale of the market can increase the demand for the specialized labor needed to create ideas. For given amount of intellectual property protection, this serves to drive up the wages of that kind of labor. We show that these increased economic rents do not serve a useful economic purpose, because they do not increase the number of ideas that are produced. Decreases in the amount of intellectual property protection that merely serve to lower these rents are socially desirable, because they reduce the monopoly distortion, without reducing the production of ideas. This observation is of some importance because the argument has been advanced by lobbying organizations such as the MPAA that those most likely to suffer from a diminuation of intellectual property protection are marginal workers and not big stars. Our analysis shows that the opposite is true, and it would be poor public policy indeed to allow large monopoly distortions in order to further enrich already wealthy individuals.

The general equilibrium approach to optimal intellectual property policy emphasizes the connection between broader features of the economy and intellectual property. We illustrate this through a number of comparative static results. Some changes, such as increased real wages, may increase the optimal time limit. Others, such as technological change in the form of computers, that reduce the size of the indivisibility required to produce an idea, reduce the optimal time limit for intellectual property.

We also consider a number of extensions of the basic model. The most important is to relax the assumption that quality is neutral. As the scale of the market or the term of intellectual property protection increases, ideas that are more marginal from a private point of view are produced. If the social value of these ideas declines even faster, then the argument for decreasing the length of protection with market scale is strengthened. On the other hand, if ideas that are more marginal from a private perspective are more beneficial from a social perspective, then a system of exclusive rights is an extremely poor method of encouraging the production of ideas: it leads to the production of the least socially useful ideas rather than the most socially useful ones. In this case a system of public prizes based on even crude estimates of social benefit would prove much superior to a patent or copyright system.

Finally, we consider the issue of "harmonization," studying countries that independently set time limits on innovation. Because some of the benefits of a higher time limit are received by the other country, when there are two countries of equal size, there is a tendency to set protection too low, and there is a "harmonization" argument to be made for international treaties raising the time limit. However, this argument applies only to two countries of equal size. When the countries are of unequal size, smaller countries tends to set low limits and free ride off of the large country - but the large country tends to set limits that are too high because it does not account for the social benefit of innovation to the smaller countries. In this case "harmonization" does not mean setting limits equal to or larger than those in the large and more protected country, but rather adjusting the time limits to lie in between the larger protection of the larger country and the smaller protection of the smaller countries.

There is an existing literature on the optimal length of patent protection. This literature, stemming from the paper of Gilbert and Shapiro [1990], examines the trade-off between patent length and bread for a single innovation. Gilbert and Shapiro give assumptions under which optimal length is infinite, while Gallini [1992] shows that with a more realistic model of the competition that surrounds breadth, this result may be reversed. This literature does not examine the broader question of optimal policy that covers many different ideas, and takes as given that policy may easily determine "breadth" as well as "length." We think that "breadth" is much more difficult to legislate than "length," and because it is less visible, more subject to rent-seeking, regardless of legislative intent. In our model, unlike this literature, we effectively take the "breadth" of protection as exogenous and focus on length. Insofar as "breadth" as well as "length" can be legislated, our parameter may be regarded as kind of a summary of length and breadth combined. Hence, it would be good public policy to reduce breadth as the scale of the market increases as well as length. In particular, it might be a good idea to introduce the "independent invention" defense as suggested by Scotchmer [2002], or to eliminate product patents in favor of process patents only. However, it is clearly more practical to tie a time limit to the growth of the economy than a particular scope of coverage.

Implicit in our formulation is that indivisibilities are individually small relative to the size of the economy. The two largest indivisible research and development efforts we are familiar with are the Manhattan Project and the NASA moon landings. [1] The cost of the Manhattan Project during 1942-1945, in 1996 dollars, is estimated at \$20 billion or approximately \$7 billion per year. GDP in 1944 and 1945 were approximately \$1700 billion per year in 1996 dollars, so the Manhattan Project cost approximately 0.4%

[1]Data for both GDP data and Manhattan Project costs are from http://www.virtualology.com/MANHATTANPROJECT.COM/costs.manhattanproject.net/; NASA budget data are from http://history.nasa.gov/SP-4102/ch7.htm

of GDP per year. NASA expenditures in 1964-68, in current dollars, were approximately \$5 billion per year. GDP in 1966 in current dollars was approximately \$790 billion, so the moon landings cost approximately 0.6% of GDP per year. Notice that both of these innovations were publicly financed. For private innovation and creation, the most costly movie production prior to the *Lord of the Rings* was *The Titanic,* which cost \$200 million in 1997,[2] while DiMasi et al [1991] report the estimated average cost of bringing a new drug to market, including the cost of failed attempts, to be \$231 million 1987 dollars. These are an order of magnitude less costly than the large publicly financed projects; they run on the order of 1/10,000 of US GDP.

## 2. The Model

Ideas are indexed by their characteristics $\omega \in \Omega$. The space of characteristics $\Omega$ is a topological measure space; for concreteness it can be taken to be a subset of a finite dimensional Euclidean space, although this plays no role in the subsequent analysis. Each idea requires a minimum amount of labor $h(\omega) \geq 0$ to be produced, created or invented, where $h(\omega)$ is a continuous function. We will often refer to $h(\omega)$ as the *indivisibility* or the minimum size for inventing/producing an idea. There can be many ideas with given characteristics; we describe this by a measure $\eta(\omega)$ representing the"number" of ideas with characteristics $\omega$ in an economy of unit size.

There is also a continuum population of agents of size $\lambda$; with the parameter $\lambda$ we aim at measuring the scale of the economy. The number of available ideas may depend on the size of the economy, so the total number of ideas with characteristics $\omega$ available in an economy of size $\lambda$ is $\eta(\omega)g(\lambda)$. To capture the principle that in a larger population more ideas of a given quality are available $g(\lambda)$ is assumed non-decreasing in $\lambda$; we may assume without loss of generality that $g(1) = 1$.

For any particular type of idea $\omega$, the amount of labor input $y(\omega)$ involved in its instantiation must overcome the indivisibility $h(\omega)$. If the input of labor is below the threshold $h(\omega)$ no prototype will emerge and no consumption is possible. Once the threshold is reached, the idea may be reproduced costlessly without limit. Let $x(\omega)$ denote consumption of an idea with characteristics $\omega$. If $y(\omega) < h(\omega)$ then $x(\omega) = 0$; if $y(\omega) \geq h(\omega)$ then $x(\omega) \geq 0$. It is convenient also to measure per capita units of consumption as $z(\omega) = x(\omega)/\lambda$.

The utility of a representative individual has a Dixit-Stiglitz form over goods of different quality. When consuming $z$ units of a good with characteristics $\omega$, consumers receive a utility of $v(z,\omega)$, where $v(z,\omega) \geq 0$ is assumed continuous in $\omega$, and in $z$ non-decreasing, and at least up to a limit $z^*$,

[2] http://history.sandiego.edu/gen/filmnotes/costs-movies.html

smooth and strictly increasing. We also assume $\lim_{z\to\infty} v(z,\omega) = v^C(\omega) < \infty$ and, to simplify the analysis of qualities that are not produced, $v(0,\omega) = 0$. Note that since $v(z,\omega)$ is bounded, $zv_z(z,\omega) \to 0$ as $z \to \infty$, that is per capita revenue falls to zero as per capita consumption grows without bound. We also assume that $zv_z(z,\omega)$ has a unique maximum at $z^M(\omega)$. Apart from consumption of idea-goods, consumers receive a utility of $\ell$ from leisure $0 \le \ell \le L$, where $L$ is the individual endowment of time; $L$ has two alternative uses: leisure or production. Leisure here is a shorthand for all activities that take place outside of the idea sector, so it includes the production of goods that do not involve ideas, or goods that use ideas that already exist. Since $\eta(\omega)g(\lambda)$ of type $\omega$ ideas are potentially available, individual utility is

$$g(\lambda)\int v(z(\omega),\omega)\eta(d\omega)+\ell.$$

Note that the marginal utility of leisure is normalized to one. The social feasibility constraint is that the amount of leisure consumed equals the amount left over after the production of prototypes.

$$\lambda(L-\ell) = g(\lambda)\int y(\omega)\eta(d\omega).$$

It is straightforward to note that profit maximization and efficiency require $y(\omega) = h(\omega)$ for all ideas that are produced, and $y(\omega) = 0$, otherwise.

**Patent Equilibrium.** Our notion of equilibrium is that of a *patent equilibrium* in which there is a fixed common length of patent protection for all ideas, as it is the case in almost all countries around the world. This means that in terms of average present value of the flow of consumption, a fraction $\phi$ occurs under monopoly, and a fraction $(1-\phi)$ occurs under competition. We refer to $\phi$ as the level of protection. We assume that there are potentially many individuals who can produce or make use of any particular idea; certainly the number of individuals who have historically had truly unique ideas is miniscule. We do not explicitly model the "patent race" by which patent is awarded, and simply assume that through some procedure, a particular individual is awarded a "patent" for a particular idea. This provides the individual with a complete monopoly over that particular good while the patent is in effect. During this period the individual chooses the production of prototypes and final output to maximize profits. Under patent protection, our economy is similar to the traditional Dixit-Stiglitz "monopolistic competition" economy. Once the patent expires, anyone who wishes to do so may make copies of the ideas that had been previously introduced under the patent regime. The patent holder loses control over the reproduction process, so for those ideas for which the indivisibility has already been overcome, competition sets in, output and consumption jump to infinity and prices fall to zero. We have assumed, mainly for the sake of simplicity, that revenue also falls to zero, although as pointed out in Boldrin and Levine [1999], this need not be, and in general it is not, the case. A type of good is produced if, given the patent length $\phi$, the prospective monopolist finds it profitable to overcome the indivisibility. This notion of equilibrium is closely connected to that of of Hart [1979], Makowski [1980], and Acemoglu and Zilibotti [1996].

The market for innovation is equilibrated through the wage rate of labor $w$. The higher is $w$ the costlier it is to produce new ideas and, therefore, fewer of them will be produced. If the amount of labor used in the production of ideas is strictly less than the total endowment $\lambda L$, wages $w = 1$. Otherwise, $w$ must be chosen to reduce demand for labor to the point where the amount of leisure is 0.

A monopolist for a good with characteristics $\omega$ who sells $z$ units of output to each of the $\lambda$ consumers receives the price $v_z(z,\omega)$. Revenue is $\lambda z(\omega)v_z(z(\omega),\omega)$, and is assumed to have a unique maximum at $z^M(\omega)$; the cost faced by the monopolist is $wh(\omega)$. For a commodity with characteristics $\omega$, $\rho(\omega) = z^M(\omega)v_z(z^M(\omega),\omega)/h(\omega)$ expresses the ratio of its private value to the innovation cost. In fact $\rho(\omega)/w$ represents one plus the rate of return on investment which would accrue to the inventor of commodity $\omega$ if patents lasted forever. We refer to $\rho(\omega)$ as the *private value* of $\omega$. The monopolist receives a fraction $\phi$ of the private value, times the size $\lambda$ of the market. Hence, he introduces the good if $\phi\lambda\rho(\omega)h(\omega) \geq wh(\omega)$, or equivalently if the wage rate satisfies

$$w \leq \phi\lambda\rho(\omega).$$

Given parameters $\lambda$ and $\phi$ and the equilibrium wage $w$ this determines a cutoff level $\underline{\rho} = w/\phi\lambda$ such that no goods with private value lower than $\underline{\rho}$ will be introduced in the patent equilibrium and all ideas with a $\rho(\omega)$ above $\underline{\rho}$ will be produced. Notice that $\underline{\rho}$ is strictly decreasing in $\phi\lambda$, meaning that as the scale of the market or the extent of protection increase, ideas with a lower private value are introduced into production. Notice also that, in general, there need not be any monotone relation between the private value $\rho(\omega)$ of an idea and it social value (to be defined momentarily); hence ideas of high social value may be introduced only for high values of $\lambda$, or even never at all, if their private value $\rho(\omega)$ is particularly low.

Per capita social welfare in a patent equilibrium is derived by integrating utility for those goods that are produced less the cost of producing them

$$g(\lambda)\int_{\rho(\omega)\geq\underline{\rho}}[\phi v(z^M(\omega),\omega) + (1-\phi)v^C(\omega) - h(\omega)/\lambda]\eta(d\omega) + L$$

This can be reformulated in a useful way in terms of $\rho$. First, observe that the measure $h(\omega)\eta(\omega)$ represents the quantity of labor socially needed to produce ideas of a certain kind $\omega$. Consider the measure $h(\omega)\eta(\omega)$, which we refer to as the *mass of ideas*, restricted to the $\sigma$-subalgebra of the Borel sets of $\Omega$ generated by the subsets of $\Omega$ on which $\rho(\omega)$ is constant; make the regularity assumption that it can be represented by a continuous density function $\mu(\rho) = \int_{\rho(\omega)=\rho} h(\omega)\eta(d\omega)$. For any function $f(\omega)$ we may define a conditional value $\overline{f}(\rho)$ in much the same way as a conditional expectation is defined. Specifically, $\overline{f}(\rho)$ is defined, $\mu$-almost everywhere, by the condition that $\int_B \overline{f}(\rho)\mu(\rho)d\rho = \int_B f(\omega)h(\omega)\eta(d\omega)$ for every set $B$ in the $\sigma$-subalgebra of the Borel sets of $\Omega$ on which $\rho(\omega)$ is constant.

Let $\nu^M(\omega) \equiv v(z^M(\omega),\omega)/h(\omega)$ and $\nu^C(\omega) \equiv v^C(\omega)/h(\omega)$ be, respectively, the *social value* of a commodity of type $\omega$ under monopoly and under competition. Reformulating in terms of $\rho$ the two social values $\overline{\nu}^M(\omega)$ and $\overline{\nu}^C(\omega)$, we may rewrite per capita social welfare as

$$g(\lambda)\int_{\underline{\rho}}^{\infty}[\phi\overline{\nu}^M(\rho)+(1-\phi)\overline{\nu}^C(\rho)-1/\lambda]\mu(\rho)d\rho+L.$$

It is useful to rewrite this using relative utility measures. Set $\sigma(\rho) \equiv \overline{\nu}^M(\rho)/\rho > 0$ to be the ratio of social value to private value under monopoly, and set $\Delta(\rho) \equiv (\overline{\nu}^C(\rho)-\overline{\nu}^M(\rho))/\overline{\nu}^M(\rho)$ to be the distortion introduced by monopoly; if $\Delta(\rho)=0$ then there is no social loss from monopoly, so we assume that $\Delta(\rho)>0$. Finally, we may then write per capital social welfare as a function of private value

$$g(\lambda)\int_{\underline{\rho}}^{\infty}[\phi\rho\sigma(\rho)+(1-\phi)\rho\sigma(\rho)(1+\Delta(\rho))-1/\lambda]\mu(\rho)d\rho+L.$$

It is useful to state our base assumptions directly in terms of these constructs. We assume that for $\underline{\rho}>0$

$$L^D = g(\lambda)\int_{\underline{\rho}}^{\infty}\mu(\rho)d\rho<\infty,$$

so that the amount of labor required to produce all ideas exceeding a particular private value threshold is finite. Notice that $\rho\mu(\rho)/w$ is the total revenue of a monopolist investing in ideas of private value equal to $\rho$. Define

$$M(\rho)=\int_{\rho}^{\infty}\rho'\mu(\rho')d\rho',$$

then $M(\rho)/w$ is the per capita monopoly revenue aggregated over all firms with private value of $\rho$ or greater. Making the regularity assumption that $M$ is differentiable, we may define $\Upsilon(\rho)\equiv-\rho M'(\rho)/M(\rho)$ as the *elasticity of total monopoly revenue* with respect to variations in the marginal private value of the ideas being introduced. We also make the regularity assumption that $\Upsilon(\rho)$ is differentiable.

To illustrate, consider the case in which utility has the quadratic form $v(\omega,z) = b(\omega)\left(Z(\omega)^2 - [z - Z(\omega)]^2\right)$ for $z \leq Z(\omega)$ and $v(\omega,z) = b(\omega)Z(\omega)^2$for $z > Z(\omega)$; that is, demand is linear. In this case we have $\sigma(\rho) = 1.5$ and $\Delta(\rho) = 0.5$ independently of the private value of the underlying goods. More generally, we can define the notion of *quality neutrality.* We have three different measures of the quality of an idea. The private value $\rho$ is one measure; the social value of monopoly output $\overline{v}^M(\rho)$ is a second measure, and the social value of competitive output $\overline{v}^C(\rho)$ is a third. By *quality neutrality* we mean that all three of these measures are proportional; that is $\sigma(\rho) = \sigma$ and $\Delta(\rho) = \Delta$ so that the social value ratio and monopoly distortion do not depend on product quality. We will consider both the neutral and non-neutral cases below.

## 3. Quality Neutrality

We first examine the case of quality neutrality, and examine how socially optimal protection $\hat{\phi}(\lambda)$ depends on market size. If the elasticity of total monopoly revenue is reasonably well behaved near $\rho = 0$ then for large enough $\lambda$ socially optimal protection must be declining with $\lambda$. If the elasticity of total monopoly revenue is increasing with $\rho$, a condition that is satisfied in most examples, then socially optimal protection is in fact non-increasing as a function of $\lambda$. As we shall see in examples, the "typical case" is that for small $\lambda$ optimal protection is $\hat{\phi} = 1$, then, once a threshold value of $\lambda$ is reached, $\hat{\phi}(\lambda)$ strictly declines. Roughly, there are two cases. If the elasticity of total monopoly revenue is increasing with $\rho$ and $\hat{\phi}(\lambda) < 1$ we can show from the first order conditions and implicit function theorem that $\hat{\phi}(\lambda)$ is strictly declining. On the other hand, if near zero the elasticity of total monopoly revenue is decreasing with $\rho$, then labor demand must be growing faster than labor supply, and so the labor constraint must bind. We also show that whenever the labor constraint binds, it must be the case that $\hat{\phi}(\lambda)$ is strictly decreasing.

**Proposition.** *Suppose quality is neutral. If for some* $\tilde{\rho}$ *and* $0 < \rho < \tilde{\rho}$, $\Upsilon'(\rho) \neq 0$ *then there exists* $\overline{\lambda}$ *such that* $\hat{\phi}(\lambda)$ *is unique and strictly decreasing for* $\lambda > \overline{\lambda}$. *If* $\Upsilon'(1/\lambda\widehat{\phi}(\lambda)) > 0$ *then* $\hat{\phi}(\lambda)$ *is unique and non-decreasing.*

*Proof.* We begin by analyzing the case in which the labor constraint does not bind, so $w = 1$. Differentiating with respect to $\phi$ and dividing out the irrelevant constant $g(\lambda)$ we get the first order condition for a social optimum.

$$
\begin{aligned}
&FOC(\lambda,\phi) = \\
&\quad [(1/\phi)\{\phi\sigma + (1-\phi)\sigma(1+\Delta)\} - 1]\,(1/\lambda^2\phi^2)\mu(1/\phi\lambda) \\
&\quad - \int_{1/\phi\lambda}^{\infty} \rho\sigma\Delta\mu(\rho)d\rho \\
&\quad = -\,[(1/\phi)\{\phi\sigma + (1-\phi)\sigma(1+\Delta)\} - 1]\,(1/\lambda\phi)M'(1/\phi\lambda) - \sigma\Delta M(1/\phi\lambda)
\end{aligned}
$$

We may divide through by $M(1/\phi\lambda)$. Since this is strictly positive, the resulting expression

$$NOC(\lambda,\phi) = [(1/\phi)\{\phi\sigma + (1-\phi)\sigma(1+\Delta)\} - 1]\,\Upsilon(1/\lambda\phi) - \sigma\Delta$$

has the same qualitative properties as $FOC(\lambda,\phi)$: it has the same zeroes, the same sign on the boundary and $NOC_\phi(\lambda,\phi) < 0$ is sufficient for a zero to be a local maximum.

We next differentiate with respect to $\phi$ to find the second order condition for a social optimum

$$
\begin{aligned}
&NOC_\phi = \\
&\quad -\,[(1/\phi)\{\phi\sigma + (1-\phi)\sigma(1+\Delta)\} - 1]\,(1/\lambda\phi^2)\Upsilon'(1/\lambda\phi) \\
&\quad - \frac{\sigma(1+\Delta)}{\phi^2}\Upsilon(1/\lambda\phi)
\end{aligned}
$$

The second terms is unambiguously negative. The first term has two factors of interest. We have $(1/\phi)\{\phi\sigma + (1-\phi)\sigma(1+\Delta)\} - 1$ representing social surplus of the marginal idea produced; since privately it yields zero profit, it must yield positive social surplus. If the other factor $\Upsilon'(1/\lambda\phi) > 0$ then there is a unique solution to the social optimization problem; if $NOC(\lambda,1) \geq 0$ then that solution is $\hat{\phi}(\lambda) = 1$, otherwise it is the unique solution to the first order condition $NOC(\lambda,\phi) = 0$.

In the latter case, we may use the implicit function theorem to compute $\hat{\phi}'(\lambda)$.

$$
\begin{aligned}
&\frac{d\phi}{d\lambda} = -\frac{NOC_\lambda}{NOC_\phi} \propto NOC_\lambda = \\
&\qquad -\,[(1/\phi)\{\phi\sigma + (1-\phi)\sigma(1+\Delta)\} - 1]\,(1/\lambda^2\phi)\Upsilon'(1/\lambda\phi)
\end{aligned}
$$

which is negative if and only if $\Upsilon'(1/\lambda\phi) > 0$. This covers the second half of the proposition when the labor constraint does not bind.

If the labor constraint does bind, increasing $\phi$ does not change the set of ideas that are produced, which is determined by the labor supply, but merely serves to increase the monopoly distortion and drive up the wage rate. Hence, if the social optimum is to allow the labor constraint to bind, $\phi$ must be chosen as small as possible subject to the constraint of full labor utilization and $w = 1$. Consequently concavity of welfare in the interior implies a unique optimal choice of $\phi$. This establishes a unique optimal policy function $\hat{\phi}(\lambda)$.

Finally, we turn to the first half of the proposition. For fixed $\phi$ and all large enough $\lambda$ we may assume that either $\Upsilon'(1/\lambda\phi) > 0$ or $\Upsilon'(1/\lambda\phi) < 0$. In either case, $\Upsilon(\rho)$ must have a (possibly infinite) limit as $\rho \to 0$. Observe that $\Upsilon(\rho) \equiv -\rho M'(\rho)/M(\rho)$, and that $M(\rho)$ is non-increasing. Suppose first that $-\rho M'(\rho)$ does not converge to infinity. If it is bounded away from zero, $M(\rho) \to \infty$, implying $\Upsilon(0) = 0$. If it is not bounded away from zero, since $M(\rho)$ is bounded away from zero, again, $\Upsilon(0) = 0$. Hence, either $-\rho M'(\rho) \to \infty$ or $\Upsilon(\rho) = 0$. The latter case implies $\Upsilon'(1/\lambda\phi) > 0$, so fix $\phi = 1$ and examine

$$NOC(\lambda, 1) = [\sigma - 1]\Upsilon(1/\lambda) - \sigma\Delta.$$

Since $\Upsilon(0) = 0$ for $\lambda$ sufficiently large $NOC(\lambda, 1) < 0$ implying $\hat{\phi}(\lambda) < 1$.

Finaly, then, suppose $-\rho M'(\rho) \to \infty$. The demand for labor is

$$L^D = g(\lambda)\int_{1/\phi\lambda}^{\infty} \mu(\rho)d\rho.$$

Differentiating with respect to $\lambda$ yields

$$D_\lambda L^D = g'(\lambda)\int_{1/\phi\lambda}^{\infty} \mu(\rho)d\rho + g(\lambda)(1/\phi\lambda^2)\mu(1/\phi\lambda).$$

Labor supply is $\lambda L$ so if $D_\lambda L^D \geq L + \varepsilon$ for all sufficiently large $\lambda$ the labor constraint must eventually bind. But $-\rho M'(\rho) = \rho^2\mu(\rho) \to \infty$ as $\rho \to 0$ so, for $\phi$ bounded away from zero, $D_\lambda L^D \to \infty$, so in this case the labor constraint must bind. □

One interesting case is the one in which $g(\lambda)/\lambda$ is increasing in $\lambda$, that is the number of available ideas increases more rapidly than the scale of the economy.[3] In this case, the labor constraint must eventually bind, and when it does the marginal private value of the marginal idea produced will increase rather than decrease: resources will be focused on producing ideas with high rather than low private value. The reader should not be mislead by the language and notice two important implications of this hypothesis: (i) while this may not necessarily mean that ideas with higher social values are also being produced in the patent equilibrium, it certainly implies that,

[3]This may be interpreted as a reduced form version of the popular claim that, in the production of ideas, aggregate increasing returns due to interpersonal externalities are the dominant force.

(ii) patent protection is less and less needed as privately more valuable ideas require less not more IP protection to be introduced. Also in this case, then, the optimal length of protection $\phi$ is strictly decreasing in $\lambda$.

The fact that intellectual property protection can drive up the wage rate for the relevant supply of highly skilled labor when the labor constraint is binding, is a very significant one in advanced economies. Lobbyist groups, such as the MPAA often point, for example, to the high cost of producing movies as a reason for strong copyright protection. Examination of the balance sheets of movies production companies shows that much of this high cost is due to the cost of paying a few "stars" large amounts of money. Since the opportunity cost of these people is generally quite small - Harrison Ford worked as a carpenter before becoming an actor, and Lars Ulrich as a service station attendant - an important effect of reducing copyright protection will simply be to lower the rents earned by these movie stars, and consequently reduce the cost of producing movies of a given quality. This is ironic, in light of the MPAA ads that feature marginal workers in the movie industry concerned about losing their jobs due to piracy. In fact, the marginal workers are paid close to their opportunity cost, and so stand to lose little through reduced copyright protection. The big stars that the ads claim will be unaffected, stand to lose much more. A similar, even if admittedly less straightforward, argument can be applied to the drug industry with respect to the wages of medical researchers in relation to that of production workers. In general, it is hard to think of a good public policy argument for promulgating socially costly monopolies in order to further enrich already rich individuals.

Other, policy relevant, comparative static results also follow from the previous analysis. Any policy or technological change increasing the marginal cost of the skilled labor needed to introduce new ideas is equivalent to decreasing $\lambda$, so generally such changes increase the socially optimal length of protection. This observation is particularly relevant when confronting policies that increase the protection from foreign labor competition (for example, by restricting immigration, or penalizing outsourcing) of selected groups of highly skilled workers: increased protection in the labor market induces, via the rationale of fostering innovation, additional protection in the product market, generating a rather vicious rent-seeking circle. Conversely, technological improvements (such as the increasing power and reduced cost of computers, or the reconstruction of the DNA code) which reduce the size of the initial indivisibility $h$ are tantamount to increasing $\lambda$, hence should engender a reduction of the socially optimal length of protection. Notice that in the case of copyright, modern computer technology has enormously reduced the indivisibility, making it possible, for example, to provide the equivalent of a multimillion dollar recording studio of 15 years ago for 10's of thousands of dollars.

**An Example.** To better understand how $\widehat{\phi}$ depends on $\lambda$ it is useful to examine the exact solution in a parametric example where $\mu(\rho) = \rho^{-\alpha}$ for $\rho \leq \overline{\rho}$, and $\mu(\rho) = 0$ for larger $\rho$. In this example the optimal length of protection is always non-increasing.

In this parametric case we may write the first order condition for a socially optimal $\phi$ as

$$\begin{aligned} FOC &= [(\sigma/\phi)(1+\Delta) - \sigma\Delta - 1](1/\phi\lambda)^{2-\alpha} \\ &\quad + (\sigma/(\alpha-2))(\overline{\rho}^{2-\alpha} - (1/\lambda\phi)^{2-\alpha}) \\ &= [\sigma/(\lambda\phi)^{2-\alpha}][(1+\Delta)/\phi - \Delta - (\alpha-1)/(\alpha-2)] + \sigma\overline{\rho}^{2-\alpha}/(\alpha-2) \end{aligned}$$

We observe that if the first order condition is satisfied and $\alpha > 2$, then we must have

$$((1+\Delta)/\phi - \Delta - (\alpha-1)/(\alpha-2)) < 0.$$

This implies that $FOC_\lambda < 0$, which, since at an optimum $FOC_\phi \leq 0$, in turn implies that $d\phi/d\lambda < 0$.

Let $FOC < 0$ at $\phi = 1$, which is the case, for $\alpha > 2$, whenever $\lambda > 1/\overline{\rho}$ holds, that is, the size of the economy is large enough. Optimality, then, always implies $\phi < 1$. This shows that the size of the economy and the upper bound on the achievable private value of ideas play a very similar role in our model; a small economy where very highly profitable ideas are available is equivalent, from the viewpoint of optimal protection, to a large one in which only not-so-profitable ideas are available.

Now, let $\overline{\rho}\lambda \leq 1$ hold, so that $FOC > 0$ at $\phi = 1$, and $\phi = 1$ is optimal. This is an economy either of very small size or in which the private value of new ideas is very low. In this case even complete monopoly cannot help as $\overline{\rho}\lambda \leq 1$ and $\phi = 1$ imply that $\underline{\rho} < \overline{\rho}$ hold, that is, no idea is ever implemented.

Finally, notice that, in general, labor demand is

$$L^D = g(\lambda)\int_{1/\lambda\phi}^{\overline{\rho}} \rho^{-\alpha} d\rho = \frac{g(\lambda)}{\alpha-1}[(\phi\lambda)^{\alpha-1} - \overline{\rho}^{\alpha-1}]$$

while labor supply is $L^S = L\lambda$. We know from before that the maximum level of $\phi$ which might possibly be optimal, call it $\overline{\phi}$, is the one at which $L^D = L^S$ and the labor constraint is binding. This gives

$$\overline{\phi} = \left[\left(\frac{1}{\lambda\overline{\rho}}\right)^{\alpha-1} + \frac{(\alpha-1)L}{g(\lambda)\lambda^{\alpha-2}}\right]^{1/\alpha-1}$$

which is strictly decreasing in $\lambda$ and $\overline{\rho}$.

## 4. QUALITY NONNEUTRALITY

In this section we drop the quality neutrality assumption and study the optimal degree of protection when private and social values are not in a constant proportion to each other, independently of $\omega$. Note that, in terms of private value under monopoly, social value under monopoly is $\sigma(\rho)\rho$ while social value under competition is $(1+\Delta(\rho))\sigma(\rho)\rho$ so that, for given $\phi$, expected social value is $\phi\rho\sigma(\rho)+(1-\phi)\rho\sigma(\rho)(1+\Delta(\rho)) = \rho\sigma(\rho)[1+(1-\phi)\Delta(\rho)]$. The two polar cases in which private and social values are, respectively, positively and negatively related seem worth considering.

If goods with lower private value have also lower social value, in the sense that $\sigma'(\rho) > 0$ and/or $\Delta'(\rho) > 0$, common sense and simple calculations show that this is a further reason for the length of protection to decline with the scale of the market. On the other hand, if either $\sigma'(\rho) < 0$, or $\Delta'(\rho) < 0$, or both, then this weakens the connection between the scale of the market and the declining optimal protection. In particular, in this case, it becomes possible to have the optimal degree of IP protection increasing with market scale, even when the mass of ideas does not increase rapidly with quality, so that the labor constraint never binds.

On the other hand, while $\sigma'(\rho), \Delta'(\rho) < 0$ may seem like an argument for increasing protection with the scale of the market, in fact $\sigma'(\rho), \Delta'(\rho) < 0$ weakens the overall case for protection. That is, $\sigma'(\rho)$, $\Delta'(\rho) < 0$ means that private return is poorly correlated with public benefit. In the extreme case, there may actually be a negative correlation between private and public benefit. In this case, the private sector produces the ideas of least social merit first, so it is certainly not worthy of IP protection when the scale of the market is small as this would simply lead to production of ideas of little social value. The argument for strengthening protection as the scale of the market increases, is that the increased scale of the market eventually leads to private sector to produce ideas that do have some significant social value, and at this point, we can try to compensate for the weakness of private incentives by increasing the level of protection. While this is formally correct, it clearly is a lopsided argument when it comes to designing welfare-improving policies. It takes as given the policy instrument, patents, even if the latter is the least adequate to maximize social welfare (because social value moves opposite to private value.) If it were really the case in practice that privately valuable innovations have little or no social value, and viceversa, then a form of government intervention other than IP would be recommendable, such as publicly sponsored research projects, or auctioning of production rights, or subsidies for innovators producing the socially valuable ideas, for example. Patents, certainly not.

**Alternatives To Government Grants of Monopoly.** The latter remarks, that when private and social values of new ideas are not aligned IP protection is the least appropriate policy instrument, suggest we should, albeit briefly, consider how alternative forms of government intervention fare in our environment. An obvious alternative to having the government award private monopolies is to have the government award prizes for innovation. This can be financed in much the same way that private monopolies raise money - by imposing a sales tax on sales of new goods. Unlike the award of a private monopoly, the tax rate does not need to be set so high as to give the monopoly revenue, and Gilbert and Shapiro [1990] show in effect that having a low tax throughout the life ot the good is preferable to having a high tax (monopoly) for part of the life of a good and low tax (after the copyright/patent expires) for the remaining life of the goods. Hence, such a system of taxes, is intrinsically less distortionary than awarding private monopolies. Insofar as the prize money is simply paid back to the innovator, this is essentially the same as a system of mandatory licensing, in which the holder of the private monopoly is required to sell at a government mandated prices. Systems of mandatory licensing are widely used - in copyright, for example such things as radio play of music and xeroxing of copyrighted materials are covered by mandatory licensing provisions. In the case of patent, mandatory licensing was widely used in Taiwan until they were forced to reform their patent system by the United States. So this kind of mandatory licensing represents, as we might expect, the efficiency improvement from replacing an unregulated monopoly with a regulated monopoly.

However, there is little reason that the proceeds of taxes on new goods should be paid back to the innovator. From an efficiency perspective, it is better that the proceeds be used to defray the costs of producing innovations of high social value. This has several advantages over an intellectual property system. First, to minimize the monopoly/tax distortion, the minimum necessary to get innovation should be paid. In particular, it is best to pay $h(\omega)$, the indivisibility, to the innovator rather than the full social value. The intellectual property system makes little use of social knowledge of $h(\omega)$; with the exception of the non-obviousness requirement (now largely defunct) of patent law, patents and copyright base reward on social value rather than social cost. Second, as we noted above, if it is indeed the case that social value is poorly correlated with private value (the strongest case for increasing intellectual property protection as the scale of the market increases) a system of rewards based on other information about social value is likely to lead to a much better mix of innovations being produced. It is important to note that, like mandatory licensing, systems of public (and private) prizes have been widely used and are of demonstrated practicality. Historians of aviation have argued that prizes played an important role in the development of the airplane. The current X-prize has unleashed an enormous amount of innovation in aerospace technology, with the first privately funded supersonic flight taking place within the last year.

The issue, also in the context of our model, boils down to the public knowledge of the true social cost of introducing a new idea. When the latter is known, a public subsidy to innovators equal to the amount $h(\omega)$, financed by a consumption tax and followed by unconstrained competition, is easily shown to provide the least distortionary mechanism. When the information about the true cost $h(\omega)$ of innovating is private the problem appears less straightforward and worthy of further investigation.

## 5. Extensions

**Consequences of Competitive Rents.** As argued in Boldrin and Levine [1999, 2002, 2004] it is by no means true that in the absence of any IP protection profits for innovators are negligible or even zero. At each moment of time, and especially shortly after innovation just took place, a capacity constraint is present that will give rise to non-distortionary competitive rents. Most likely, there are also first-mover advantages, such as those documented by Tofuno [1989] in the market for financial securities.

We can model the presence of capacity constraints in our setting, by assuming a capacity constraint $\bar{z}(\omega)$ on per capita production of the type $\omega$ good after the IP protection expires. In this case, assuming the capacity constraint is not binding during the period in which the innovator is a legal monopolist (otherwise there would be no reason to allow any IP protection) when we take account of the competitive rent accruing to the innovator, his total revenues from production of a new idea are

$$\phi\lambda z^M(\omega)v_z(z^M(\omega),\omega)+(1-\phi)\lambda\bar{z}(\omega)v_z(\bar{z}(\omega),\omega).$$

Assuming the labor constraint does not bind, this means production takes place when revenues are at least equal to the cost of the indivisibility, $h(\omega)w$. Dividing through by $h(\omega)$, and noting that $w=1$ when the labor constraint does not bind, we may write

$$\phi\lambda\rho(\omega)+(1-\phi)\lambda\rho^C(\omega)\geq 1$$

where $\rho^C(\omega)=\bar{z}(\omega)v_z(\bar{z}(\omega),\omega)/h(\omega)$ is the competitive rent per unit of indivisibility cost. Let us use the simplifying assumption that competitive rent is proportional to monopoly revenue, per unit of indivisibility cost; that is $\rho^C(\omega)=\vartheta\rho(\omega)$, with $0<\vartheta<1$. Note that this is stronger than our earlier neutrality assumptions, which had to hold only in expected value. Then we may again write social welfare entirely in terms of $\rho$ and the only modification of our earlier expression for social welfare is that $\underline{\rho}=[(\vartheta+$

$\phi(1-\vartheta))\lambda]^{-1}$. The corresponding NOC is

$$\begin{aligned} NOC(\lambda,\phi) &= (1-\vartheta)\left[(1/(\vartheta+\phi(1-\vartheta)))\left\{\phi\sigma+(1-\phi)\sigma(1+\Delta)\right\}-1\right] \\ &\quad \times\Upsilon(1/(\vartheta+\phi(1-\vartheta))\lambda) \\ &\quad -\sigma\Delta \end{aligned}$$

from which we see that it continues to be true that if

$$\Upsilon'(1/(\vartheta+\phi(1-\vartheta))\lambda)>0$$

then $\hat{\phi}(\lambda)$ is unique and non-decreasing. In this case, we may also compute $NOC_{\vartheta}<0$, so that higher competive rents lead to a reduction in the optimal level of protection.

Finally, consider that the NOC at $\phi=0$ is

$$(1-\vartheta)\left[(1/\vartheta)\left\{\sigma(1+\Delta)\right\}-1\right]\Upsilon(1/\vartheta\lambda)-\sigma\Delta$$

, and as $\lambda\to\infty$ this approaches $-\sigma\Delta$. Hence, when the population is sufficiently large, rather than the optimal level of protection asymptoting to zero it should instead be set equal to zero at some finite market size. This result is reinforced in practice because there are many direct and rent-seeking costs of operating a system of intellectual property, and there are other first mover advantages besides the competitive rent. Ultimately, when the market has expanded enough, IP should be eliminated entirely.

**Variations on the Utility Function.** We assumed that utility is linear in the output of the idea sector and in labor. We can consider more generally, the functional form

$$\begin{aligned} &U\left(g(\lambda)\int_{\underline{\rho}}^{\infty}[\phi\rho\sigma(\rho)+(1-\phi)\rho\sigma(\rho)(1+\Delta(\rho))]\mu(\rho)d\rho\right) \\ &+V\left(L-(1/\lambda)\int_{\underline{\rho}}^{\infty}\mu(\rho)d\rho\right). \end{aligned}$$

To see how this works, we examine the sharp case in which $U$ is linearly increasing to an upper limit, then constant, while $V$ is linear all thorough its domain. In other words, once total output in the idea sector is large enough, no further utility is generated, or $U(x)=x$ for $x\leq X$ and $U(x)=X$ for $x>X$. This is a simple way of modeling the following intuitive case, which follows from the assumption that $g(\lambda)/\lambda$ is increasing in $\lambda$: marginal utility of consumption decreases faster than the marginal utility of leisure as the former goes to infinity and the latter remains bounded above by $L$. As long as $\lambda$ is small enough, satiation does not affect the original equilibrium, and the price of output in the idea sector $P=1$. However, once $\lambda$ grows large enough, the satiation constraint binds.[4] The equilibrium requires that exactly $X$ be produced in the idea sector, so the price in the idea sector $P < 1$ must fall to discourage labor from flowing into that sector. But a more efficient way to discourage labor from flowing into the idea sector is simply to lower $\phi$. This simple result is kind of the mirror image of the labor constraint binding in the previous analysis.

More general non-linearities in $U$ have a similar effect as long as our basic assumption about relative marginal utilities of consumption and leisure holds. The latter can be reformulated in the following way: as the size of the market grows and more ideas are produced, the price in the idea sector declines faster than the marginal utility of leisure for fixed $\phi$, hence the relative price of labor goes up. This seems to be reflected, in the real world, by the century-long decline in hours worked per capita. Notice that it is immaterial to our argument if the increase in the size of the market is due purely to an increase in the population, or to an increase in trade, or to unmodeled technological advances that reduce the size of the indivisibility $h(\omega)$. In general it will be best to exploit the opportunity offered by an increase in the size of the market by reducing $\phi$, rather than by allowing the relative price of labor to rise as the latter does not engender any efficiency gain but purely redistributes wealth from the general population to the skilled labor used in the production of ideas. The effect of $V$ being non-linear is similar. As leisure declines, wages rise gradually. Like the sharper effect of hitting the labor constraint, it is best to offset this by reducing $\phi$.

**Positive Marginal Cost of Distributing Ideas.** So far we have assumed that there is no marginal cost of distributing ideas and of producing goods using ideas. There are several possible cases, depending on which inputs are needed to distribute and produce ideas. One possibility is that the same labor used to produce ideas is used to distribute them and to produce goods using them. This case is rather complex, because it introduces a third margin into the choice of $\phi$ - the monopoly for inframarginal ideas; the marginal ideas, and the amount of labor used to produce existing ideas.

However, in practice the type of labor used to distribute ideas and to produce copies of ideas is probably not a terribly good substitute for the labor used to produce the ideas themselves. If we introduce an additional factor of production - unskilled labor, call it - and assume that this is used to distribute ideas and produce goods using ideas, provided that this factor is in plentiful supply, so that marginal cost is constant, little changes in our analysis. In particular, instead of examining $v(z,\omega)$, we should examine per capital utility net of the cost of distribution and production using the idea:

[4]Notice that $U'(X) < V'(L)$ is pretty much all that is needed here.

$v(z,\omega) - mc(\omega)z$. This may have an impact on whether quality is neutral, since it may be neutral for $v(z,\omega)$, but not for $v(z,\omega) - mc(\omega)z$, but, for example, with quadratic utility (linear demand) we have neutrality in both cases.

If the unskilled labor constraint binds before the skilled labor constraint - something we think is unlikely, but since the demand for unskilled labor in this model grows much more rapidly than the demand for skilled labor as $\lambda$ increases, something we recognize is a possibility - then the situation again is complicated by the existence of a third margin. Unskilled labor must be allocated between producing more of particular goods, versus producing more kinds of goods. The point is, though, that the central finding of our analysis of the basic case does not change. As the size of the market increases, if the unskilled labor constraint binds before the one for skilled labor does, the choice is between shifting more unskilled labor to activities that make use of new ideas or to activities that make use of old ideas. When the marginal social value of the latter is larger than that of the former a reduction in the level of IP protection $\phi$ is recommended.[5] When the opposite is true, and assuming again that private and social values are properly aligned, unskilled labor becomes, in this model, a rescaled version of skilled labor and the previous analysis, which recommends a reduction in $\phi$ when the size of the market increases, applies.

## 6. HARMONIZATION

We now turn to the issue of IP protection in the world economy. We assume that there are $k$ countries and that each country $i$ has a fixed fraction $\theta_i$ of world demand, labor and world ideas. The total size of the world economy is still $\lambda$.We examine the case in which countries may not discriminate against foreign inventors. While *de-facto* violated in some occasions, this reflects current legal practices around the world, and it allows us to focus on the specific role of IP protection. We let $\phi_i$ denote the level of IP protection in country $i$. Throughout this section we assume there is complete and costless free trade of goods, that the labor constraint does not bind and that the elasticity of total monopoly revenue is increasing.

From an inventor's perspective, what is relevant is the effective (weighted by market shares) total protection received world wide. This is simply $\phi = \sum_i \phi_i \theta_i$. Since the labor constraint is assumed not to bind, $\underline{\rho} = 1/\phi\lambda$ continues to determine the marginal invention. Each country is assumed to maximize own social welfare, which is given by

[5] Also in this case, as before, should the private and social values move in opposite direction the analysis acquire a further degree of complication.

$$g(\lambda)\int_{\underline{\rho}}^{\infty}[\phi_i\rho\sigma+(1-\phi_i)\rho\sigma(1+\Delta)-1/\lambda]\mu(\rho)d\rho+L,$$

the significant feature of which is that the monopoly distortion depends on $\phi_i$and not on $\phi$. Note that in this world, each country is assumed to get a *pro rata* share $\theta_i$ of the total mass of world innovations, so that net revenue to a country from selling/purchasing IP-protected products on the world market is zero.

The NOC corresponding to the social welfare problem is

$$NOC_k(\phi_i,\phi_{-i})=\theta_i\,[(1/\phi)\,\{\phi_i\sigma+(1-\phi_i)\sigma(1+\Delta)\}-1]\,\Upsilon(1/\phi\lambda)-\sigma\Delta$$

Because the elasticity of total monopoly revenue is assumed increasing, this is strictly concave in $\phi_i$ and continuous as a function of $(\phi_i,\phi_{-i})$ so has a pure strategy Nash equilibrium, characterized by the first order conditions $NOC(\phi_i,\phi_{-i})=0$.

Consider first a symmetric equilibrium of a symmetric model in which $\theta_i=1/k$. In equilibrium we must have $\phi=\phi_i$, or

$$NOC_k(\phi)=(1/k)\,[(1/\phi)\,\{\phi\sigma+(1-\phi)\sigma(1+\Delta)\}-1]\,\Upsilon(1/\phi\lambda)-\sigma\Delta=0.$$

Let $\phi^1$be the solution to the single country problem and $\phi^k$the symmetric solution to the symmetric $k$ country problem. Because the first term in $NOC_k(\phi)$ is positive and $\theta_i<1$ we have that $NOC_k(\phi^1)<0$, implying, since $NOC_{k\phi}<0$ under the elasticity condition, that $\phi^k<\phi^1$. Notice that $\phi^k$ is decreasing in $k$ so that, as the number of countries increases the symmetric Nash equilibrium converges to the case of no IP, which is suboptimal in our setting. The intuition behind this result is simple: by decreasing $\phi^k$ a country loses because it creates fewer new goods and gains because it consumes at the competitive level the goods created by the remaining $k-1$ countries. As $k$ increases the second margin strictly dominates the first.

The concavity of the objective function in $\phi_i$ implies that each country choosing the solution to the single country problem is in fact the unique social optimum for the world as a whole. Moroever, if each country is constrained to set the same level of protection as all others, for example through a legal mechanism such as the WTO, they would all agree to choose the social optimum $\phi^1$. This is the standard harmonization result: in the unconstrained protection game countries underprotect duo to the public goods nature of protection, and a WTO-like mechanism that forces harmonization leads them to the second best.

Unfortunately, this analysis, while mathematically correct, has little relevance to policy analyis for two reason. First, it assumes that countries choose policies to maximize their own social welfare. The evidence of vastly increased levels of protection in response to vastly increased market size provides relatively conclusive evidence that this is not how countries choose their levels of protection; rather it suggests that levels of protection are determined largely by rent-seeking. The fact that copyright extensions have applied retroactively is yet another piece of evidence that the level of protection is driven by rent-seeking and not social efficiency. Second, current extensions of IP are not between countries of equal size with currently equal levels of IP, who by harmonizing will agree to a common higher and more efficient level of protection. Rather, they are between large countries with high levels of protection and small countries with no protection. To study this, the relevant model is not one of $k$ identical countries, but rather of one large country (the U.S. and the E.U.) and $k-1$ identical small countries.

Consider then, a situation where there is one large country with share $\theta_1$and $k-1$ small countries with shares $\theta_i = (1-\theta_1)/(k-1) < \theta_1$. First we show that, in this case, regardless of $k$ the equilibrium level of protection $\phi$ is bounded away from zero. The NOC for the large country is

$$NOC_1(\phi_1,\phi) = \theta_1\left[(1/\phi)\left\{\phi_1\sigma + (1-\phi_1)\sigma(1+\Delta)\right\} - 1\right]\Upsilon(1/\phi\lambda) - \sigma\Delta = 0.$$

Observe that $\phi \geq \theta_1\phi_1$ and recall that NOC is decreasing in $\phi$. Hence, $NOC(\phi_1,\theta_1\phi_1) \geq 0$. Since this latter expression is also decreasing in $\phi_1$ a solution to $NOC(\widetilde{\phi}_1,\theta_1\widetilde{\phi}_1) = 0$ must satisfy $\phi_1 \geq \widetilde{\phi}_1 > 0$. This in turn implies that in equilibrium $\phi \geq \theta_1\widetilde{\phi}_1 > 0$. This shows that $\phi$ is bounded away from zero independent of $k$ because the large country will never impose a negligible amount of protection.

We now turn to the NOC for the small countries. At $\phi_i = 0$ this is

$$\begin{aligned} NOC_k(0,\phi) &= (1-\theta_1)/(k-1)\left[(1/\phi)\left\{\sigma(1+\Delta)\right\} - 1\right]\Upsilon(1/\phi\lambda) - \sigma\Delta \\ &\leq (1-\theta_1)/(k-1)\left[(1/\theta_1\widetilde{\phi}_1)\left\{\sigma(1+\Delta)\right\} - 1\right]\Upsilon(1/\theta_1\widetilde{\phi}_1\lambda) - \sigma\Delta \end{aligned}$$

which is strictly negative for $k$ larger than

$$K = \frac{\sigma\Delta}{(1-\theta_1)\left[(1/\theta_1\widetilde{\phi}_1)\left\{\sigma(1+\Delta)\right\} - 1\right]\Upsilon(1/\theta_1\widetilde{\phi}_1\lambda)} + 1.$$

Since there is always a unique solution to $NOC_k(\phi_i,\phi) = 0$, for $k > K$ it occurs at $\phi_i = 0$. In this case the solution for the large country is simply the solution to the social optimum problem which ignores demand and supply from the rest of the world; that is, the choice of $\phi_1$is the one that would be optimal for a population of $\theta_1\lambda$. By the usual scale of market effect, that means the equilibrium solution for $\phi_1$is larger than the value that maximizes world social welfare, that is, the solution to the social optimum problem with population $\lambda$.

In short, in the empirically relevant case where there are small countries with no IP and large countries with excesive IP, effective harmonization requires not only that all countries be constrained to set the same level of IP protection, but that the level of IP protection chosen be substantially less than the existing levels of protection in the large countries.

Notice, incidentally, that our base assumption is that countries cannot increase the level of domestic innovation by changing their IP laws, as their share of total innovation is fixed at $\theta_i$. Insofar as countries can increase their share of world-wide innovative production by changing their national level of IP protection, they benefit from the fact that they get a disproportionate share of the total revenue from innovation. In fact there is some evidence that favorable IP treatment can attract innovation. There are several reasons for this. First, favorable IP legislation may be a signal of favorable treatment of innovators in general (for example, as in Ireland, through tax law). Second, although legal discrimination against foreign inventors is forbidden in principle, there may be a variety of informal reasons why it is advantageous to be a domestic innovator to take advantage of strong local IP protection. Finally, the distribution of innovation across countries can be driven by the explicit rent-seeking behavior of innovators, who may choose to reward countries that provide favorable IP protection with increased revenue from domestic innovation. The movie industry, for example, has gone to some length to reward and punish its political foes with movie production.

Insofar as increasing IP protection lures innovation, a second type of equilibrium distortion arises. Rather than underprotecting in an effort to free ride off of innovation in other countries, the incentive is to overprotect to try to get a disproportionate share of IP revenue.

## 7. CONCLUSION

The most important missing aspect of our analysis is the dynamic feature that ideas build on other ideas. As pointed out in Boldrin and Levine [1999, 2003] and Scotchmer [1991], ideas that use other ideas as input greatly weakens the case for IP because the latter, while it encourages innovations by improving the return to the first inventor, it discourages further innovations through raising their cost. In this sense, there is no reason to think that adding dynamic features to the model is likely to make IP more socially desirable. In fact, when the complexity of innovations increases because new ones need to use more and more old ideas as inputs, the presence of widespread IP naturally determines a hold-up problem where even one residual monopolist may prevent new ideas from being implemented.

Turning to our results: if $\phi\lambda$ is held fixed, the quality of ideas produced remains unchanged. This is not the social optimum: generally we will want to take advantage of the increased $\lambda$ to allow some marginal ideas to enter the market. A simple rule of thumb that allows for some additional marginal ideas to enter is to reduce the actual length of term in proportion to $\lambda$. Because of discounting, halving the length of protection will reduce $\phi$ by less than half. Thus, the simple rule of thumb would be that if the size of market doubles, the length of protection should be halved. For example, the G7 nations account for about 2/3rds of world GDP. If we think of the intellectual property changes in the WTO as extending the protection that exists in the G7 to the rest of the world, this suggests a reduction in the length of term by about 1/3rd. Similarly, if the world economy is growing at 2% a year, a simple rule of thumb would be to reduce protection terms by 2% per year. A paradigmatic case is that of popular music. Forty years ago, at the time of Elvis Presley and The Beatles, new recordings selling a million units were considered exceptional successes and awarded "golden records" while in the current times a successful record sells easily ten or tweny millions copies. The effective size of the market has, therefore, increased of at least a factor of ten. At the same time, advances in recording and digital technologies have reduced the fixed cost required to produce a new record to about one fifth of its earlier level. This suggests that the socially optimal length of copyright protection should have dropped of about a factor of fifty. Unfortunately, in the case of copyright, terms have been moving in the opposite direction; copyright terms have grown by a factor of about four since early in the twentieth century. This means that, at least for recorded music, they currently are two hundred times longer than they should be. A similar calculation can be performed for books and movies. Consider the fact that, since the beginning of the past century, world GDP has grown by nearly two orders of magnitude. It is reasonable to argue that the size of the market for books and movies must have grown of grown of at least as much, as literacy has surged and the availability of playing devices has increased more than proportionally due to the dramatic drop in their relative prices. Hence, if the copyright term of 28 years at the beginning of the 20th century was socially optimal, the current term should be about 3 months, rather than the current term of approximately 100 years. This gives a ratio of four hundred between the actual copyright terms and their socially optimal value. The fact that copyright terms have risen so drastically, while at the same time the scale of the market has expanded even more drastically is, we think, strong evidence that copyright terms are not set to reflect the solution to a second best social optimization problem, but rather in response to rent-seeking behavior by copyright holders.

## REFERENCES


[1] Acemoglu, A. and F. Zilibotti (1996), “Was Prometheus Unbound by Chance? Risk, Diversification and Growth, *Journal of Political Economy,* **105**, 709-751.

[2] Boldrin, M. and D. K. Levine (1999), “Perfectly Competitive Innovation,” University of Minnesota and UCLA, November.

[3] Boldrin, M. and D.K. Levine (2002), “The Case Against Intellectual Property,*”The American Economic Review (Papers and Proceedings)* **92**, 209-212.

[4] Boldrin, M. and D.K. Levine (2003), "Rent Seeking and Innovation," *Journal of Monetary Economics*,

[5] Boldrin, M. and D.K. Levine (2004), “IER Lawrence Klein Lecture: The Case Against Intellectual Monopoly,*” The International Economic Review,* vol (May) .

[6] DiMasi, J. A., R. W. Hansen, H. G. Grabowski, L. Lasagna (1991), “The Cost of Innovation in the Pharmaceutical Industry,” *Journal of Health Economics,* **10,** 107-142.

[7] Gallini, N. (1992), “Patent Policy and Costly Imitation,” *Rand Journal*, **23,** 52-63.

[8] Gilbert, R. and C. Shapiro (1990), “Optimal Patent Length and Breadth,” *Rand Journal*, **21,** 106-112.

[9] Hart, O. D. (1979), “On Shareholder Unanimity in Large Stock Market Economies,” *Econometrica,* **47,** 1057-83.

[10] Makowski, L. (1980), “Perfect Competition, the Profit Criterion and the Organization of Economic Activity,” *Journal of Economic Theory,* **22**, 222-42.

[11] Maurer, S. M. and S. Scotchmer (2002), “The Independent Invention Defense in Intellectual Property,” *Economica,* forthcoming.

[12] Scotchmer, S. (1991), “Standing on the Shoulders of Giants: Cumulative Research and the Patent Law,” *Journal of Economic Perspectives,* **5.**