Review of Economic Research on Copyright Issues, 2005, vol. 2(1), pp. 45-67

INTELLECTUAL PROPERTY AND THE EFFICIENT ALLOCATION OF SOCIAL SURPLUS FROM CREATION

MICHELE BOLDRIN AND DAVID K. LEVINE

Abstract. In the modern theory of innovation, monopoly plays a crucial role both as a cause and an effect of creative economic activity. Innovative firms, it is argued, would have insufficient incentive to innovate should the prospect of monopoly power not be present. This theme of monopoly runs throughout the theory of growth, international trade, and industrial organization. We argue that monopoly is neither needed for, nor a necessary consequence of innovation. In particular, intellectual property is not necessary for, and may hurt more than help, innovation and growth. We show that, in most circumstances, competitive rents allow creative individuls to appropriate a large enough share of the social surplus generated by their innovations to compensate for their opportunity cost. We also show that, as the number of pre-existing and IP protected ideas needed for an innovation increases, the equilibrium outcome under the IP regime is one of decreasing probability of innovation, while this is not the case without IP. Finally, we provide various examples of how competitive markets for innovative products would work in the absence of IP and critically discuss a number of common fallacies in the previous literature.

1. Introduction

We discuss positive and normative issues related to the appropriation of the surplus generated by creative activity. We take for granted that an act of creation generates a positive total net surplus, i.e. that the total opportunity cost of the inputs is less than the social value of the output, and try to understand: (i) How such surplus is allocated between creators, imitators, and consumers as a function of two different systems of property rights, and of the consequently different market structures; (ii) Which, among the two allocations of surplus, maximizes social welfare and, hence, which system of property rights and which market structure are socially desirable. We label the two property right systems considered, respectively, IP and NIP.

Under IP, as is currently the case in most advanced countries, legislation attributes to the creator of an idea the right to impose substantial, and essentially unlimited, restrictions on the usage that lawful buyers of copies of the idea are allowed to undertake after the purchase takes place; in particular, no one can make further copies of the idea without permission from the creator. Under NIP, creators own the copies of their idea they either produce or acquire, but cannot impose restrictions on the usage that lawful buyers of copies of their ideas can undertake; in particular, buyers of copies are allowed to make further copies by using existing technologies. Under the IP system of property rights, the market structure is monopolistic and the creator is, de facto, an eternal (for copyrightable materials) or an almost eternal (for patents) monopolist. Under the NIP system of property rights, the market structure may be monopolistic in the very few initial periods, if there is a unique creator who is then a natural monopolist, but becomes competitive in the sequel. In our modeling exercise, in order to allow for the frequent and important case of multiple simultaneous creators, we will assume competition from the very first period under NIP, and consider the particular case in which the creator is initially a natural monopolist as an extension.

The analytical part of the paper contains the following items. We introduce a model of innovation based on a set of assumptions that, as argued in the example of the next Section, are the fundamental characterizing features of creative activity. For this model, we derive the equilibrium under the NIP system and characterize its properties, including the case in which the creator is a natural monopolist in the first period. After briefly mentioning a few extensions, we compare the NIP allocation with the one that obtains in the IP system, under the same assumption for preferences and technology. Next we revisit the standard model of innovation, in which no innovation takes place under NIP. While these may not be, and are not in our view, the empirically relevant circumstances, they are those almost invariably adopted to argue for the necessity of strong IP legislation. Even in these special circumstances, our analysis casts abundant doubts on the established wisdom, according to which more IP protection is, from a social viewpoint, always better than less. Further on, we show by means of a simple example that in a paradigmatic (and everyday more frequent) case the presence of IP protection may eventually lead to an ever decreasing amount of creation; the paradigmatic case consisting of a situation in which the creator of a new idea must use previous ideas, controlled by previous creators/monopolists, in the inventive process. Finally, we use our model of creative activity to illustrate how "superstars" arise when the technology to reproduce ideas becomes more efficient. This allows us to show, even if this should not be the concern of a policy aimed at maximizing social welfare, that a NIP system would not necessarily "impoverish" creative individuals, but instead it would allow the best among them to still earn large amounts of money. In fact, we show, incomes disproportionally larger than those of other (slightly less talented) creators accrue in the NIP equilibrium to the best creators. In the IP allocation, this "disproportion" between relative income and relative talent only increases, making income inequality larger without gains for social welfare. Before summing up we revert to the "purely verbal" mode and discuss a number of common fallacies often encountered in the literature supporting strong IP regimes. In the conclusion we summarize our findings, debate a few other controversial points, and claim that an almost complete abolition of IP protection would lead to more creative activity, a more efficient allocation of its surplus, and higher social welfare.

1.1. **Relation with Previous Literature.** What we refer to as the "standard model" assumes that innovations involve a fixed cost of creation and a constant marginal cost of reproduction, that the new idea or creation is not embodied in anything the creator owns and controls or, at least, that it can be imitated and copied by third parties at no cost other than the marginal cost of reproduction, without additional revenue accruing to the original creator because of this copying activity. In other words, the standard model assumes that, once invented, the idea is tantamount to a public good upon which the original innovator would have no control if IP protection were absent. In these circumstances the creator could only sell the copies he produces at a price equal to the marginal cost of reproduction, without any ability of earning additional rents and, thereby, without the possibility of recovering the original fixed cost. This model is so widespread and so frequently used that it is hard to pinpoint its exact origin; some passages in Arrow (1962) may have provided it a theoretical foundation, even if it is probably more appropriate to attribute the basic idea that "monopoly is good, indeed necessary, for innovation," to the Schumpeter of *Capitalism, Socialism and Democracy*. A classical treatment, arguing that the current (at the time) IP system is the optimal one, can be found in Nordhaus (1969), while Romer (1986) and the new growth literature that followed it are certainly the most recent and successful among its theoretical developments.

Without doubt we are not the first to question the appropriateness of the IP regime for fostering creation and innovation even if, to the best of our knowledge, no one had argued that it applies in general and to most innovations and, again to the best of our knowledge, no formal dynamic general equilibrium theory of competitive innovation had been submitted before Boldrin and Levine (1999). We hasten to add, though, that we view our work as a straightforward application to the issue of innovative activity of classical capital theory and, in particular, of the Marshallian theory of competitive rents under limited capacity. The fundamental point we make is that classical competitive theory applies to ideas, and their copies, the same way it applies to wheat, and its copies. Plant (1934) and Stigler (1956) are two illustruous predecessors, who argued the same point verbally, but not less forcefully, a long time before us; Stigler (1956), in particular, contains a penetrating and convincing discussion of competitive innovation that should be mandatory reading for every researcher in the field.

In more recent years, a number of authors have argued, more or less informally or on the basis of empirical research, that ease of copying by third parties does not necessarily prevent creators of literary or artistic works from collecting substantial revenue. This point has been made, for example, in the literature on copying (Besen and Kirby (1989), Liebowitz (1985), Johnson (1985)), and sharing (Ordover and Willig (1978)). An excellent and updated survey of this and related literature, cum critical discussion, can be found in Legros (2005). The basic message of these articles is, nevertheless, different from the one advocated here, both in the formal setup and in the substantive conclusions. Liebowitz (1985), for example, carries out an econometric study of the impact of photocopying on the revenue of journal publishers, to conclude, on the basis of his empirical estimates, that through means such "as indirect appropriability, exposure effects, and price discrimination" publishers' revenue may actually have increased due to the advent of photocopying. In all these works, the model adopted, or the underlying framework of reasoning in the case of empirical work, is the standard one, with fixed costs, increasing returns to scale, and a monopolistic innovator who can actively set prices and discriminate between different buyers. Further, there is no modeling of the dynamic of creation and reproduction, of the decision problem faced by potential innovators, and of the way in which different market structures determine different allocations of the surplus and, hence, different incentives to create. These are the issues we address in our work.

2. Embodiment of Ideas and Surplus Appropriation

Before moving on to the mathematical part, we illustrate verbally the basic intuitions underlying our analysis, using a well-known historical example to fix ideas.

Economic, and more generally social, progress is the long run, and altogether surprising, result of the continuous creation of new commodities, of their free exchange among individuals, and of the competition among producers of different goods, be they creators or imitators. Economists have long realized that there would be but a slow and possibly inconsequential improvement in human living standards without sustained innovation. This point was argued, most forcefully, by Joseph Schumpeter in *The Theory of Economic Development* (1911). With constant technology and a constant set of goods, the process of capital accumulation, when based only on the saving of a share of the yearly income flow, would generate but a fraction of the growth in per capita income we have witnessed since the inception of human history. Accumulation of capital under a constant technology, history and common sense conjure to suggest, cannot go very far due to the presence of fixed resources and the diminishing returns they bring about. Innovation is the engine of change and economic development, hence understanding its nature, internal mechanisms, and the social and institutional factors that bring it about or impede it, is, we believe, the single most important problem faced by the social sciences. It is our contention that understanding innovation is tantamount to understanding competition, that the latter is a necessary condition for the former and that, under very general circumstances, it is also sufficient. If innovation is the flow that enriches us all, then competition is the spring from which it erupts.

Innovation, for us, is the creation of the first copy of a good/process/idea that did not exist before. As the word "idea" will be used henceforth to denote all innovations, its usage should be briefly clarified. In our terminology, Isaac Netwon's innovation did not consist just in "thinking" the gravitational laws, but in the process of embodying them in his mind first, and in formulas and written expositions later. When, in 1687, he completed the manuscript of his *Philosophiae Naturalis Principia Mathematica* and had it published, "Newton's innovation" was completed. All subsequent copies of the *Principia* were reproductions of that first copy of his idea, and they were produced with a technology different from the one he had to use to obtain his first manuscript. Notice, that with "copy" here we refer here to either a physical copy of the actual book or the (equally physical, if less visible) copy of the gravitational laws embodied in the brain of another scientist or layman, i.e. a piece of socially valuable human capital. Indeed, and this is something crucial, the social value of Newton's innovation is more properly measured by the number of copies of his laws existing in the second form (actual human capital) than in the first (copies of the book.) All such copies stemmed from Newton's original copy and the social value of the latter would have been much smaller, or even negligible, without them. Newton's reward, either in terms of intellectual prestige or in terms of actual wealth and social status, became so high because very many copies (of either type) of the *Principia* were eventually reproduced. In our terminology, the first copy of the gravitational laws is the "prototype" and it embodies, for the first time, Newton's idea; the *innovation technology* is the one Newton adopted to figure out the gravitational laws and write the *Principia*. The *imitation technology* is the one used by subsequent publishers of the book and by whoever learned and understood the content of the *Principia*. Notice, which is relevant, that the *Principia* were published before the Statute of Anne introduced some (weak by current standard) degree of IP legislation in the U.K.

Notice that the final products of the two technologies are, functionally speaking, equivalent: a copy of the *Principia* is a copy of the *Principia*, and a human that understands the principles and laws of gravitation is, at least from this narrow point of view, equivalent to any other human who understands the same principles and laws. This point will become relevant later on, when discussing the *public domain* for ideas. Notice also that both technologies use a variety of inputs to obtain their final product, that some of these inputs are previous innovations (e.g. Kepler's Laws) and that such inputs can be acquired on competitive markets under NIP, but would have to be obtained from monopolists by acquiring many licenses under IP. There are two exceptions, to what we just said. First, the innovation technology uses a particularly scarce input, Newton's geniality in this case, which greatly limits the number of initial prototypes that can be obtained. Had we been concerned with a less dramatic invention, simultaneous creation by a number of different and independent innovators would have been likely, as it is often the case in practice. Still, the total amount of "creative ability" available at any point in time to make prototypes of a new ides is quite limited. In the jargon of economics, there is always *limited creative capacity* of prototypes at any given point in time. In the particular case of scientific inventions or of artistic creation, this limitation of creative capacity may persist for a long time: new scientific discoveries are very difficult to understand, that is why we have Ph.D. programs and post-docs, and live performances of, say, new hard to imitate music, which is why live concerts are often sold out and very expensive. The imitation technology also uses a special kind of input, and that is a pre-existing copy of the *Principia* (in case we are considering a publisher making copies of the book) or, generally, someone who has already understood its content (in case we are considering a student learning gravitational laws). Either way, this particular input(s) is also in limited supply; strictly speaking, this is true at any point in time and even now, but it is especially true in periods close to the time in which the first prototype of the *Principia* appeared. In summary, the imitation technology also faces a *limited productive capacity*, the size of which is basically determined by the number of copies of the idea "Newton's Gravitational Laws" embodied in humans/books at any point in time.

A little reflection shows that this set of properties is not specific to the particular case of the *Principia*, but applies quite widely (we would say: universally) to other innovations. The differences are quantitative, never qualitative: new valuable ideas are always embodied in either people or things; innovative capacity is always limited; imitation/reproduction always requires copies of the idea and hence stems from the original prototype even if in some rare cases imitation may not require large investments; reproductive capacity is also quite limited for a substantial number of periods after the innovation takes place; new ideas almost always require old ideas to be created, and creation is more and more a complex and cumulative incremental process; finally, consumers are always impatient and would rather have the stuff today than tomorrow. Our theoretical analysis builds upon such properties, and an additional one: it took quite a while to Newton to come up with the gravitational laws (falling apples notwithstanding) and, for what we know, even longer to fully articulate them in the manuscript of the *Principia*. Further, the *Principia* were not a minor, infinitesimal departure from or improvement upon previous knowledge, but a substantial one indeed. This property is also general, at least qualitatively!

Producing the prototype, via the invention technology, requires quite often a large investment, which we want to think of as an *indivisibility*. While it is not true that a sizeable indivisibility is involved with the production of prototypes of every idea, it is true that this is often the case, and that this feature of creative activity should be taken in proper account when discussing the allocation of economic surplus from creative activity.

Finally, a few words to further clarify our approach to the problem. We ask what is socially optimal, and how incentives should be provided (i.e. which market structure can provide the appropriate incentives) for the *socially optimal amount of creative activity to take place*. The problem of providing incentives for innovation should not be confused with the protection of rents of intermediaries, or rents of established artists, or creators more generally. The issue here is not what makes creators richer or as rich as possible, but how to allocate to them enough of the surplus from creative activity so that they have the incentive to carry it out efficiently, from a social view point. This requires focusing on the concept of *opportunity cost*, i.e. to ask: when a potential innovator considers the choice between engaging in creative activity or doing something else, his opportunity cost is determined by how much income he would receive from doing something else. Efficiency requires that, should the innovator opt for creation, he receive from the latter at least as much as he would receive from the alternative activity, that is: his opportunity cost. When the market structures allows the innovator to receive more than his opportunity cost, this additional rent serves no socially useful purpose. Per se, this additional rent may just be a pure transfer, which does not affect economic efficiency; nevertheless, more often than not, and in particular when monopoly power is involved, this additional rent accrues to the innovator because he has the incentive to provide less innovations, or less copies of his innovations, than socially efficient. In this case the additional rent is not just a neutral transfer from consumers to innovators (which may be unfair, but irrelevant for efficiency) but a socially costly and inefficient tax on consumers, less copies of ideas are available to the people than it is desirable and technologically feasible. Our critique of current IP laws focuses mainly on this second aspect.

Technological innovation continuously change the opportunity cost and reservation values of the various agents involved in creation. So, for example, the invention of the printing press made the craftsmanship accumulated over century of artisans and monks unnecessary for copying or for production of new books. This was a blessing, for writers of books and their readers, but also a curse for those artisans who suddenly lost their long established title to a substantial share of the social value of every book, new or old that it be. Given current technologies, and the continuous improvement in the innovation and reproduction technologies, it would be crucial to measure what the opportunity costs of creators and innovators actually is. Unfortunately, this is an endeavor to which applied economists, especially in the area of industrial organization, have dedicated minimal attention and we are not aware of any study estimating the minimum future expected income needed to attract potential innovators into creative activity.

3. A Model of Creation

3.1. **Creation under Perfect Competition.** To understand whether an innovation will take place or not in a NIP (or competitive) environment, we must understand how much a new idea is worth *after* it is created. Consider a competitive environment in which some innovation has already been produced. In other words, there are currently some prototypes, books, songs, or blueprints owned by the creator. There may be many independent creators, as simultaneous inventions are allowed by our theory. To simplify, and without loss of generality, assume there is a continuum of identical creators, all of which behave competitively in the market for the new good, i.e., in analogy with traditional growth models, consider the case of a "representative competitive innovator." Again for simplicity, focus on the extreme case where every subsequent item produced using the template is a perfect substitute for the template itself – that is, what is socially valuable about the invention is entirely embodied in the product. At a moment in time, each item has two alternative uses: it may be consumed or it may be used to produce additional copies. For simplicity we assume that while the process of copying is time consuming, there is no other cost of producing copies beside the cost of (partially) withdrawing from consumption the goods that are used as templates in the reproduction process. The analysis that follows makes it transparent that adding a vector x of other, competitively marketed, inputs to the reproduction process only complicates the algebra without altering the conclusions.

Specifically, suppose that currently the innovators have made $k > 0$ units of the new product available. These units can be sold on a competitive markets, and purchasers can either consume them or invest them as inputs to make further copies. Obviously, innovators can also participate as purchasers in the market, as long as they remain price takers on either side of the market. Suppose that $0 \leq c \leq k$ units are allocated to consumption, leaving $k - c$ units available for the production of copies. The $k - c$ units that are copied result in $\beta(k - c)$ copies available in the following period, where $\beta > 1$. Because the units of the good used in consumption might be durable, there are ζc additional units available next period. In many cases $\zeta < 1$ due to depreciation, however we allow the possibility that the good may be reproduced while being consumed, and require only that $\zeta < \beta$, reproducing while consuming is not easier than just reproducing.

Besides the representative innovator, there is also an infinitely lived representative consumer, who receives a utility of $u(c_t)$ from consumption in each period, where u is strictly increasing, concave, and bounded below, and who has a discount factor $0 < \delta < 1$. We assume that the technology and preferences are such that feasible utility is bounded above. Given the initial stock of prototypes k_0 this defines a simple model of capital accumulation under competition in the presence of a linear technology (again, linearity is not essential.) It is well known that competitive equilibria are Pareto optima under these conditions, hence the equilibrium allocation solves the planner's problem

$$\max \sum_{t=0}^{\infty} \delta^t u(c_t)$$

subject to

$$k_{t+1} \leq \beta(k_t - c_t) + \zeta c_t.$$

It is well known that the solution to this optimization problem may be characterized by a concave value function $v(k_0)$, which is the unique solution of

$$v(k) = \max \{u(c) + \delta v(\beta k - (\beta - \zeta)c)\}$$

subject to $0 \leq c \leq k$.

In an infinite horizon setting, beginning with the initial stock of the new good $k_0 = k$ we may use this program recursively to compute the optimal k_t for all subsequent t. Moreover, the solution of this problem may be decentralized as a competitive equilibrium, in which the price of consumption services in period t is given by $p_t = u'(c_t)$. From the resource constraint

$$c_t = \frac{\beta k_t - k_{t+1}}{\beta - \zeta}$$

If ζ is large enough relative to β it may be optimal not to invest at all and to reproduce solely by consuming. We first take the case where consumption is strictly less than capital in every period. By standard dynamic programming arguments, the price q_t of the durable good k_t can be computed as

$$q_t = v'(k_t) = p_t \frac{\beta}{\beta - \zeta}$$

This is positive and, when either the current value of consuming the idea is very high (high p_t), or the reproduction rate is very high, or the durability of the good embodying the idea very close to the reproduction rate, the NIP price of a copy of the idea is also very high. Notice that q_0 is, under NIP, the total earnings accruing to the initial innovator(s) for each prototype they sell in the, supposedly competitive, market for prototypes. The zero profit condition, also implied by NIP and competition, implies that q_t decreases at a rate of $\frac{1}{\beta}$ per period of time, hence ideas that are hard to reproduce sell at substantial prices for a long time.

This already proves the main point: innovators, even in the presence of simultaneous innovations by many independent agents, can earn positive rents when selling their new product under conditions of perfect competition. This is due, crucially, to the fact that it takes time to produce additional copies of the new good, and that the initial k_0 copies determine a capacity constraint on feasible consumption in the first and in all subsequent periods. This provides the foundation for a theory of competitive innovation under NIP. It is readily apparent that this is true as long as at least one of the following two assumptions holds, (i) there is a delay in production (i.e., consumers are impatient), (ii) capacity is bounded at all points in time (i.e., $k_0 < \infty$ and $\beta < \infty$.) What remains to be checked, and will be discussed momentarily, is how the parameters of the model determine the size of the rents accruing to the competitive innovators. Notice that, while the innovators receive positive rents, consumers are also receiving surplus in the form of the difference between the utility of their consumption flow, and the cost of renting (or buying) the good, $p_t c_t$ (or $q_0 k_0$). Furthermore, the first welfare theorem implies this is the socially efficient allocation of the surplus between innovators, imitators and consumers.

3.2. **The Decision to Create.** Consider, then, the problem faced by a competitive innovator *before* the innovative effort is undertaken. After the innovation has occurred, the representative innovator has, say, k units of the new product that he must sell into a competitive market. In a competitive market the initial units sell for $q_0 k = v'(k)k$, which may be interpreted as the rent accruing to the fixed factor owned by the innovative entrepreneur(s). The market value of the innovation corresponds, therefore, to the market value of the first unit of the new product. This equals, in turn, the net discounted value of the future stream of consumption services it generates. Introducing those first k units of the new good entails some cost $C > 0$ for the innovator, say in terms of leisure foregone while inventing or because some other goods need to be purchased to carry out the innovation process that leads to the prototypes. Consequently, the innovation will be produced if and only if the cost of creating the innovation is less than or equal to the rent resulting from the innovation and captured by the fixed factor, $C \leq q_0 k$.

The reader may suspect that something dubious hides behind the fact that we are not modeling explicitly the process through which the prototypes are created, and how the amount k is chosen. This is not the case. Imagine that, in the period before $t = 0$ the representative innovator is endowed with some resources, say L units of time, which can be used in the production of a basket of existing goods and earn a wage w. Let the production function of the prototypes be $k = I(l)$ for $l \geq \underline{l}$, and $k = 0$ for $l < \underline{l}$, with $\underline{l} \leq L$ representing the indivisibility mentioned in section 2. Taking the market value q_0 of the initial copies as given, a competitive innovator maximizes own profit by solving

$$\max_{0 \leq l \leq L} \delta q_0 I(l) - wl$$

Then, as long as $\delta q_0 I'(\underline{l}) \geq w$ holds, the representative innovator sets $\delta q_0 I'(l^*) = w$ and brings $k = I(l^*)$ prototypes of the new good to the market, at a (sunk) cost of $C = wl^*$. The case in which $\delta q_0 I'(\underline{l}) \geq w$ does not hold, i.e. the indivisibility is binding, is important and will be considered next. Obviously, this is just one particular formulation of the R&D production function. An alternative, and appealing, one assumes there is a continuum of heterogeneous innovators $i \in \left[0, \overline{I}\right]$, each innovator can produce at most one prototype at an opportunity cost of w_i, with the latter increasing in i. In this case, the "marginal innovator" I^* is determined by the equality $\delta v'(I^*) = w_{I^*}$. In general, as long as the expected C is smaller than the expected revenue $\delta q_0 k$ from selling the prototypes, competitive innovation makes perfect sense. This should also explain our insistence, in the previous section, on the dramatic need for empirical studies determining what the opportunity cost of innovators would be under NIP and competition.

A less obvious question is: What happens as β, the rate at which copies can be made, increases? If, for example, the advent of the Internet makes it possible to put vastly more copies than in the past in the hands of consumers in any given time interval, what would happen to innovations? Observe that $\frac{\partial q_0}{\partial p_0} > 0$ and that

$$\frac{dq_0}{d\beta} = u''(c_0)\frac{dc_0}{d\beta} - u'(c_0)\frac{\zeta}{(\beta - \zeta)^2}$$

When β is sufficiently large relative to ζ the first term will dominate. For concreteness, consider the case of full depreciation, $\zeta = 0$. In this case the rent will increase if initial-period consumption falls with β and will decrease if it rises. In other words, the relevant question is whether consumptions are substitutes or complements between time periods. If they are substitutes, then increasing β lowers the cost of consuming after the first period and causes first period consumption to decline to take advantage of the reduced cost of copies in subsequent periods. This will increase the rent to the fixed factor and improve the chances that the innovation will take place. Conversely, if there is complementarity in consumption between periods, the reduced cost in subsequent periods will increase first-period consumption of the product and lower the rent.

3.3. **Elasticity of Demand.** It is instructive to consider the simple, even if extreme, case in which the utility function has the CES form $u(c) = -\frac{1}{\theta c^{-\theta}}$, $\theta > -1$. In this case, it is possible to explicitly compute the optimal consumption/production plan. Consider first the case of inelastic demand, where $\theta > 0$. Here there is little substitutability between periods and a calculation shows that as $\beta \to \infty$ initial consumption increases toward an upper bound, which is nevertheless smaller than the initial stock of prototypes. Consequently, rents from innovation fall, but not toward zero. Competitive innovation still takes place if $\underline{p}\overline{c} = u'(\overline{c})\overline{c} > C$.

More relevant is the case of elastic demand, where $\theta \in (-1, 0)$. This implies a high elasticity of intertemporal substitution in consumption. In this case, utility becomes unbounded above as β increases toward a high, but finite, value. The latter details is a particular feature of the CES function form we are considering, and is not essential. What is essential, and general, is that with elastic demand revenues to the creator increase as β increases. The case of elastic demand is especially significant, because it runs so strongly against conventional wisdom: as the rate of reproduction increases, the competitive rents increase, despite the fact that over time many more copies of the new good are reproduced and distributed. The basic assumption is simply that demand for the new product is elastic.

Many people may find the idea that demand for new ideas is, at least initially, elastic not as "realistic" and "relevant" as we think it is. Indeed, some academic commentators of our work, e.g. Klein, Lerner and Murphy (2002), have made exactly this point. These critics quickly dismiss our assumption as "unrealistic" pointing out that, obviously, demand cannot be elastic because, otherwise, monopolists would be lowering prices until the point of unitary elasticity is reached. Hence, they conclude, if demand were elastic inefficiency would not be an issue even under IP; further, because monopolists do not seem to be lowering prices all that much in markets for, e.g., recorded music, then demand is certainly not elastic. This criticism, even if "obvious", is actually incorrect, both as a matter of elementary theory and of publicly available data. As we all learned in college, demand elasticity alone matters only when marginal cost is zero. In the general case with positive marginal cost, what matters is whether marginal revenue is bigger than marginal cost (so the monopolist wants to expand output) or smaller (so he wants to reduce it). Hence you would expect monopolists to produce where marginal revenue is equal to marginal cost, and this provides information about demand elasticity if we know something about marginal costs. Without going too far from the case often mentioned as "dramatic" should a NIP regime be adopted, we can try to figure out what the numbers say for music CDs. For music producers it is clearly not the case that marginal cost is zero. We have fairly precise data for the European music market (and we are willing to bet the data for the US market are similar). Over continental Europe, the final consumer price is of about 19 Euros per CD, which breaks down, roughly, like this: 7.6 euros to the music producers, 7.6 euros to the retailers, 3.8 euros for the intermediary distributors. From the viewpoint of the monopolist copyright holder the sum of the last two, 11.4 euros, is marginal cost. In fact, this is a lower bound on the marginal cost, as the monopolist has also to pay the musicians some money, and has to produce and wrap the CD. Again, according to European data, this is about 1.8 euros. Hence, marginal cost (MC) is between 11.4 and 13.2, while price (P) is equal to 19. We know that, denoting with $\varepsilon < 0$ the elasticity of demand, the monopolist sets

$$\frac{P}{MC} = \frac{\varepsilon}{1+\varepsilon}$$

In the case at hand, this gives ε between -2.64 and -3.27, depending on which value you take for MC. Demand is clearly elastic, and this should not be a surprise to anyone who buys music, or books, or DVDs! A similar exercise with books will get you a similar result; we leave it to the reader to guess what the result would be if we repeated the calculations for medicines, especially the truly innovative and important ones. That is why we claim that "elastic demand" is the empirically obvious and relevant case for creative works.

3.4. **A Few Extensions.**

3.4.1. *Inelastic demand and satiation.* There is plentiful evidence that in practice the indivisibility of ideas is not more substantial than that of other commodities, for example, automobile plants or shipyards. There is also much evidence that ideas flourish in competitive markets without government intervention in the form of patents and copyright. However, for books, music and movies, it is easy to imagine that changes in computer technology that make copying cheaper and more rapid will lead to a β so large as to cause k_t to expand so rapidly as to flood the market and drive price to zero rather quickly. We just argued that this is not, indeed, very realistic and that, under elastic demand, this would not hurt the innovator. However, let us assume, for the sake of the argument, that demand for the idea becomes rapidly inelastic and that some satiation point is reached where the market price at which copies can be sold is close to zero. It is worth noting that the same technological change is reducing the cost of books, music, and movies creation as well, so that C is also decreasing quite rapidly, and this may well offset the improved copying technology. Moreover, even if we accept that the market for copies may be quickly flooded, there are still tremendous advantages in being first. We will not attempt to enumerate those all of those advantages here. In the case of innovations, secrecy is an obvious method of generating a short-term monopoly. In the case of books and movies, most sales take place within three months of initial release. So if it is possible to keep copies encrypted for even so short a period of time, substantial revenues may be realized regardless of the quality of copying technology. Overwhelming empirical evidence from the pharmaceutical industry suggests that the first mover advantage is quite substantial, be it due to reputation effects, slow information diffusion, or simply "capture" of the medical profession. In any case, the evidence shows that most generic drugs, selling at a quarter of the price and being clinically and functionally perfect substitutes for the original products, never capture more than 50% of the market (Caves et al (1991), Congressional Budget Office (1998)).

3.4.2. *First Mover Advantage.* The first mover advantage is a form of monopoly accruing to the original innovator. A monopolist, unlike a competitor, will not allow quantity k_t to expand until the satiation point (or the inelastic portion of demand) is reached, which drops the price to zero, but will restrict output to a much lower value, the one that maximizes his revenues. This case does nothing but reinforce our point: an innovator who is really unique and, therefore, a natural monopolist for the first few periods will earn even more than the competitive one studied in the earlier part of this section. Hence, for this individual, the indivisibility C will be even less binding because, when solving the problem

$$\max_{0 \leq l \leq L} \delta q_0 I(l) - wl,$$

he will not take the initial price q_0 as given, but instead recognize that $q_0 = v'(k_0 = I(l))$. As a monopolist he can manipulate q_0 by appropriately choosing l, as he is the only producer of k_0, i.e. the only supplier of the specially skilled labor l. Clearly, the initial number of prototypes chosen by the monopolist, call it $k_0^M = I(l^M)$, will be strictly less than the one chosen under competition, $k_0 = I(l^*)$, hence socially inefficient. Nevertheless, this inefficiency, apart from being unavoidable, is also small in a NIP regime, as free competition, and faster reproduction of copies, will set in right after the first period. As we show below the "legal" (should we say "un-natural"?) monopolist created by the IP regime, will instead restrict supply in all periods, keeping the price too high and the allocation away from the efficient one for the indefinite future.

3.4.3. *Indivisibility.* We have seen how under some circumstances there may be underprovision of ideas due to indivisibility. This is due to the fact that, when the indivisibility C is particularly large and the marginal utility of the new idea is rapidly decreasing $C > q_0 k$ may obtain. In this case the traditional notion of competitive equilibrium would not do, equilibrium may not even exist and the second welfare theorem may fail: there are allocations that are socially optimal and require creation to take place, but cannot be achieved by the traditional price taking arrangement of competitive equilibrium. This is a difficult problem, that economic theorists have too long ignored and about which little is known; we cannot even attempt to address it here, hence we will recognize its existence and add just a few remarks.[1] The first is empirical: how relevant is the case $C > q_0 k$ in practice? Because the competitive analsys of innovation has been happily ignored so far, no one has any reasonable idea. At least, no one has a reasonable idea based on a systematic analysis of real world data. Our prior is that this is not a very frequent case, i.e. that most innovations are of an incremental nature and, even if they imply an indivisibility, the latter is very seldom large compared to the size of the market. Casual observation suggests that very large aggregate indivisibilities occurr a lot less often than one may think. Consider the size of some very big ideas: The Manhattan Project (1942-1945) cost \$7 billion per year in 1996 USD; GDP in 1944-1945 was about \$1700 billion per year in 1996 USD, hence the Manhattan Project cost approximately 0.4% of GDP. The overall operating budget of NASA (1962-73) runs at about \$18 billion per year in 2000 USD; the Apollo project was about 1/3 of it. In 2000 USD, the GDP of the USA in 1968 was about \$3,700 billion, implying that landing on the moon cost approximately 0.15% of GDP. Notice that these are all ideas financed by the public purse, so let us now turn to the really relevant ones, those financed by private entrepreneurs under the IP regime. The famous big movies "The Titanic" and the "Lord of the Rings" cost \$200 million each in 1997 USD; DiMasi et al [1991] estimate the average cost of bringing a new drug to market at \$231 million 1987 USD, including clinical trials (which are a public good). Hence, privately financed ideas have an indivisibility that is, at most, 1/10,000 of

[1] The interested reader should consult Boldrin and Levine (1999), (2004b) and (2004c) for a more extensive discussion of its implications.

US GDP. Our own empirical analysis, see Boldrin and Levine (2004c), suggests that, at least in the case of fictional novels, the size of C, i.e. the opportunity cost of the effort required to write one novel for the marginal writer, hardly exceeds $ 60,000 in 2004 USD; a miniscule amount indeed. Finally, the proverbial cherry on the top of the pie: we are all familiar with the "Genoma Project" and the fact that, because its indivisibility was so large, it had to be financed by the public purse. Still, we are also all aware that, in spite of the public project to be already underway and, hence, in spite of the fact that the most of the (de)coding of the human genoma was going to be in the public domain, a relatively little private entrepreneur was able to pay for that large indivisibility, recover it, and make a non negligible profit by competing against the public enterprise. From this we can safely conclude that, at least for the idea called "human genoma", the marginal utility of copies must not be rapidly decreasing.

Our second remark is admittedly more speculative. It says that, while traditional price taking behavior may not work when C is too large, other competitive arrangements, such as contingent bids, contracts that mimic lotteries on the delivery of the idea, "innovation clubs", and a number of yet unexplored and nontraditional tools that creative and competitive entrepreneurs can come up with, may well do the trick in most cases. This is a line of investigation that has been very seldom pursued, hence we cannot do anything more than conjecture that it is a fruitful terrain for future research. The traditional solution to this problem is, instead, IP: the government provision of monopoly through patents and copyright. That is, by granting control over how all copies of an idea are used, the government allows the patent or copyright holder to limit reproduction and restrict supply. This increases profits, and so provides a greater incentive to create or innovate. Our investigation, instead, suggests that more IP does more harm than good.

3.4.4. *Rent seeking.* One of the key problems with government grants of monopoly is the rent-seeking it induces. That is, when governments give away monopolies, there is incentive for would-be monopolists to waste resources competing for the award. In the case of intellectual monopolies, the resources wasted by competing "would be monopolists" takes several forms. The most widely studied is the patent race, where too much effort is invested in innovating quickly in order to be the first to get the patent. Another classical problem is the effort wasted building "work alike" innovations in order to get a portion of the monopoly. This is the case, for example, in textbooks, where every textbook is just different enough from the best-seller in the field to avoid violating the copyright. It is also the case in pharmaceuticals, where more time and effort is spent developing copycat drugs to get the share of a lucrative market, than is spent developing genuinely new drugs.

One of the worst aspects of public rent seeking is the regulatory capture or "monopoly creep" it induces. In the case of regulation, it has been observed that over time the regulatory agency becomes captured by the regulated industry, and far from imposing the public interest on the industry, serves instead to enable collusion and monopolistic practices within the industry. Similarly, in the case of patents and copyrights, over time both the scope and duration of monopoly power has been increased as a consequence of constant rent seeking. The term of copyright has risen in the USA, for example, from 28 years to 95 years; and many areas of thriving innovation not traditionally subject to patents, such as business practices, are now patentable. So while in a theoretical sense, it might be desirable to have copyrights and patents lasting a few months or a few years, as a practical matter, once copyrights and patents are allowed at all, their term and scope is likely to begin to creep upwards.

The existence of public rent seeking is not to say that there is not private rent seeking as well. For example, in the absence of patents, innovators are likely to increase their reliance on trade secrecy. Indeed, one argument for patents is that it replaces trade-secrecy, and forces innovators to reveal the secrets of their inventions. Unfortunately, as anyone who has read a patent will realize, the "secret", if there is one, is rarely revealed in a useful way in the patent application. Moreover, since patents last 20 years, the only reason to get a patent is if the inventor thinks he cannot keep the secret for that long. We have studied this issue in Boldrin and Levine (2004a), showing that creating public rent seeking is not a good way to solve the problem of private rent seeking.

3.5. **Creation Under Monopoly.** What, finally, if we are in the IP regime and the initial innovator is a legal monopolist, independently of there being one or more simultaneous innovators? What, in other words, if one of the innovators is, in force of IP laws, granted patent or copyrights over the idea, and all its future copies? The response is straightforward. The un-natural monopolist will not

$$\max_{0 \le l \le L} \delta q_0 I(l) - wl,$$

as the natural monopolist did earlier on, but, instead,

$$\max_{\{l, c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \delta^t u'(c_t) c_t - wl$$

subject to the constraints that $k_0 = I(l)$, $k_{t+1} \le \beta k_t - c_t + \zeta c_t$. It is trivial to see that the solution to this problem implies: (i) the initial quantity $k_0^M < k_0$, where the latter is the competitive solution and the former the monopolistic one, hence (ii) $c_t^M < c_t$, meaning less consumption and welfare in each period. Further, under general assumptions on the utility function $u(c)$, there exists a level $c^M < \infty$ at which the monopolist maximizes his long run profits. Once this point is reached, the monopolist has no incentive to let reproduction to continue, and will use its IP power to maintain the stock of copies of ideas at the level $\beta k^M + (\zeta - 1)c^M$, instead of letting it grow unbounded as in the NIP allocation. Much less social welfare under IP than under NIP, not just in the short but also in the very long run. Further, imagine the particular idea we are considering can be used, when a sufficient stock is accumulated, to create a new idea that reduces the cost of producing additional consumption but cannot reduce the cost of increasing the current level, and so on as additional ideas are created. That is, assume we have a ladder of ever better ideas, feeding on each other and reducing the cost of increasing the production of additional units of the consumption good. The monopolist will not choose to innovate because any investment to do so must necessarily reduce current-period revenues below the maximum, while it cannot raise revenue in any future period. Similarly, the monopolist will not allow anyone else to innovate. If this reminds you of the state of our telephone service industry in the long decades *before* the ATT monopoly was broken down, you are quite correct.

4. The Standard Model Examined

Supporters of IP like to use another model of innovation, one that, in the present context, is equivalent to the extreme case in which there is not an indivisibility, but a fixed cost in the innovation technology, and in which the parameter β in the reproduction technology is equal to infinity. That is, the creator puts resources equal to C into the innovative process, generates a new idea but the later is not embodied in anything the creator can actually control, be it either her human capital, a prototype of the good, or a master record. The idea, instead, is somehow published in the daily newspaper, or showed on free access TVs, and everyone can readily learn it and copy it. Conventional wisdom in economics is that IP environments are necessary in this setting to recover fixed costs. With demand that is perfectly elastic up to an upper bound, there is no cost of monopoly, so this would seem the ideal environment to impose IP restrictions. Our goal is to point out that this is correct only if it is not possible to produce similar items – for example textbooks that are sufficiently different to be entitled to a separate copyright, but sufficiently similar as to make no difference to consumers. When there are many firms competing for monopoly rents, and market conditions are such that rents can be obtained even with some degree of competition, the rent seeking behavior of competing monopolists dissipates the social surplus by overproduction of too many similar items. IP does better than NIP when there is one firm, but not when there are many. Moreover, consider allowing consumers to submit contingent bids. This allows the recovery of fixed costs even when there is a single firm, so NIP does no worse than IP in this case. With many firms and contingent bids, NIP essentially always does better than IP.

4.1. **One Firm.** Begin by considering a firm that faces a fixed cost $F < 1$ and can produce unlimited quantities at marginal cost zero. There are many identical risk neutral consumers; at a price of one or less, taken together they demand one unit; and will purchase nothing at price higher than one. In addition consumers who make a purchase can themselves produce additional units at a marginal cost of $\varepsilon \geq 0$. This can represent the fact that consumers have access to an inferior reproduction technology, that consumers have a preference for buying from the original producer, or that in a setting where production takes time, as in section 3, the seller of the original unit can claim a possibly substantial rent.

We consider two legal environments: one in which consumers are prohibited from reselling, which we refer to as the IP environment; and one in which downstream licensing agreements are not legally enforceable, which we refer to as the NIP environment.

4.1.1. *IP environment.* We first suppose that consumers are legally prohibited from reselling units, either by an explicit law, or by enforcement of licensing agreements that prohibit resale. In this case, the monopolist will produce one unit at a price of one, covering the fixed cost and leading to a Pareto efficient outcome.

4.1.2. *NIP environment.* Next we suppose that licensing agreements that prohibit resale are not legally enforceable. In this case, competition among consumers will force the price to $p = \varepsilon$. If $\varepsilon \geq F$ the firm will produce one unit, and again the outcome will be Pareto efficient. If $\varepsilon < F$ the firm will not produce and the outcome will be inefficient.

4.1.3. *IP and NIP with contingent offers.* The contractual environments considered so far exclude the important possibility that consumers may submit contingent bids prior to production. We consider only symmetric equilibria in which all consumers submit the same contingent bid. In the copyright environment this makes no difference: the monopolist has no reason to accept contingent bids for less than 1 since he will be able to sell at a price of 1 to anyone whose bid he does not accept. In the no copyright environment in which $\varepsilon \geq F$ contingent bids make no difference for the same reason – the monopolist will be able to sell at a price of ε to anyone whose bid he does not accept, and cannot sell at a higher price. However, when $\varepsilon < F$ in the NIP environment it makes a substantial difference: if all consumers bid F for their share of the market, the good will be produced. No higher price will clear the market, since consumers could bid less, and their bid would still be accepted. However, if any consumer bids less than F no bids will be accepted and the good will not be produced, so each consumer is decisive, and this is in fact a subgame perfect equilibrium. Hence, with contingent offers, the NIP environment implements the first best. In this case the efficiency of IP and NIP are the same, although NIP creates a substantial transfer from producers to consumers.

4.2. **Many Firms.** Now suppose that there are many identical firms that produce perfect substitutes, all using the technology described above.

4.2.1. *IP environment.* Let $p(n)$ be the post-entry price when n firms have chosen to enter the market. As a simple model of post-entry competition, suppose that $p(n) = \min\{1, (1-\alpha(n))nF + \alpha(n)\}$, where $\alpha(n) > 0$. That is, the post-entry price lies between the price needed to recover costs and the monopoly price in a way that depends on the number of firms. In this case entry will occur until n is so large that $(n+1)F > 1$, while $nF \leq 1$. For convenience let us suppose that there is actually an $n = \frac{1}{F}$; then this will be the equilibrium number of firms, and the social surplus will be zero, as the benefit to consumers will equal to the cost of production. This particular form of competition results in what we can describe as the Pareto worst outcome.

4.2.2. *NIP environment.* As in the case of one producer, the price will still be ε. If $\varepsilon < F$ there will still be no output and no social surplus. Otherwise the number of firms will be such that $(n+1)F > \varepsilon$ and $nF \leq \varepsilon$. Assuming that there is actually an $n = \frac{\varepsilon}{F}$ this will be the equilibrium number of firms, and the social surplus will be $1-\varepsilon$. This gives a reversal of the one firm result: without licensing social surplus is never lower and sometimes higher in NIP than in IP.

4.2.3. *IP and NIP with contingent offers.* With contingent bids and many firms there is a coordination problem because consumer must decide which producers to submit their bids to. After bids are submitted, let us suppose that firms are ordered by the number of bids they receive, and let $b(i)$ be the number of bids received by firm i. We assume that these are fixed numbers, and that $b(i+1) < b(i)$. We assume that all consumers submit identical bids, so $\sum_i b(i) = 1$.

Consider the NIP case first. Suppose that $\varepsilon < F$, all consumers bid p where $b(1)p + (1-b(1))\varepsilon = F$. Then the first producer exactly recovers production costs by accepting all bids and selling to the remaining consumers at the price ε. No other producer can earn a profit by entering. Consumers' expected utility is exactly F since they are risk neutral, so they are willing to bid p. In addition if any consumer bids less than p the good is not produced at all, and he is strictly worse off. Finally, there is no equilibrium with a higher value of p since then each consumer could bid less and still have the bid accepted.

Next, look at the IP environment. If $b(1) = 1$ so there is no coordination problem, then it is an equilibrium for all consumers to bid F and the first best is obtained. Suppose, however, that $(n+1)F > 1$, while $nF < 1$, so that the equilibrium without contingent bids is strict. Since there can be no more than n firms producing in any equilibrium, regardless of whether firms accept or reject bids, the effect on demand is at most $nb(1)$. Consequently if $b(1)$ is small enough the equilibrium number of firms will remain at n and the equilibrium with contingent bids will be essentially the same as the equilibrium without contingent bids, and similarly inefficient.

In summary, even in this extreme and fairly unrealistic environment, and even under the strong assumptions of perfect information and identical consumers the only case in which IP does better than NIP is with a single producer and no contingent bids. With contingent bids, NIP does at most ε worse than IP, and, when there are many firms and a severe coordination problem among consumers, NIP does substantially better than IP.

4.3. **Sequential innovation.** Here we summarize a simple result, first reported in Boldrin and Levine (2005a) and then extended and fully analyzed in Boldrin and Levine (2005b). The basic intuition is much older than our modeling, as its name, "we are sitting on the shoulders of giants", reveals, and its application to the problem of optimal IP was already discussed, albeit informally, in, e.g., Scotchmer (1991).

Suppose that to create a new idea requires the use of N existing ideas. Assume the cost of producing a copy of each of these ideas is $\frac{\varepsilon}{N}$. Under IP the value of the new idea for a monopolistic creator is ρ, while under NIP a competitive creator gets only $\phi\rho$, with $0 < \phi < 1$. With NIP, there will be many copies of each of these existing ideas competing with each other, and the inventor can obtain all N of them for a total cost of ε. Without government intervention, this socially desirable invention will take place, provided only that $\phi\rho \geq \varepsilon$.

Assume IP applies to the N old ideas as well as the new, and that the owners of the old ideas only know that ρ is drawn from a uniform distribution over $[0, \overline{\rho}]$. Each sets a price p_i to license her invention. Then, if owners of all the other existing ideas are setting the price p, each owner of an existing idea receives an expected revenue of

$$\frac{\overline{\rho} - (N-1)p - p_i}{\overline{\rho}} p_i$$

If $2\varepsilon < \overline{\rho}$ the Nash equilibrium of this game is at $p = \frac{\overline{\rho}}{N+1}$, and therefore the inventor must pay $\frac{N\overline{\rho}}{N+1}$ to acquire all the N licenses. Therefore, only if the value of the new idea satisfies $\rho > \frac{N\overline{\rho}}{N+1}$ will the potential new creator be able to acquire all N licenses and move forward with creation. This occurs with probability $\frac{1}{1+N}$, implying that under IP the probability of creation goes to zero as its complexity (number of licenses required) increases. By way of contrast. under NIP the probability of innovation is $1 - \frac{\varepsilon}{\phi\rho}$, which is independent of the number of required inputs.

The moral is simple: as $N \to \infty$ the additional incentive for innovation under an intellectual property regime is more than completely offset by the additional cost it imposes on innovation. As technologies grow more and more complex, requiring more and more specialized inputs, the monopoly power induced by patents and copyright becomes more and more socially damaging.

5. Superstars

The phenomenon of superstardom was defined by Rosen (1981, p. 845) as a situation "wherein relatively small numbers of people earn enormous amounts of money and dominate the activities in which they engage." Its puzzling aspect derives from the fact that, more often than not, the perceivable extent to which a superstar is a better performer or produces a better good than the lesser members of the same trade is very tiny. Is superstardom due to some kind of monopoly power, and would it disappear in a competitive environment?

Our theory shows that when there are indivisibilities, technological advances in the reproduction of "information goods" may lead to superstardom, even under perfect competition. Hence, our model predicts that superstars should abound in industries where the main product is information, which can be cheaply reproduced and distributed on a massive scale. Such is the case for the worlds of sport, entertainment, and arts and letters, which coincides with the penetrating observations (p. 845) that motivated Rosen's original contribution.

For simplicity, we consider a world in which all consumption takes place in a single period, but our results extend directly to an intertemporal environment. There are two kinds of consumption goods. The first is the information good we concentrate upon, while the second can be interpreted as a basket of all pre-existing goods. Specifically, we assume utility of the form $u(c)+m$, where c is the information good. There are two kinds of potential producers, A and B, each with a single unit of labor. The two producers are equally skilled at producing the second good: a unit of labor produces a unit of the second good. However, A produces information goods that are of a slightly higher quality than those produced by B. To be precise, we assume that one unit of type A labor can produce $(1+\varepsilon)\beta$ units of good c, while one unit of type B labor can produce β units of good c.

This case, without indivisibility, does not admit superstars, in the sense that the price of type A labor must be exactly $1+\varepsilon$ the price of type B labor. Since type A labor is more efficient at producing the information good, type B labor will be used in the information sector only after all type A labor is fully employed in that sector. Suppose that this is the case. Let ℓ_2 denote the amount of type B labor employed in the information sector. Then, the equilibrium condition is simply $\beta u'(\beta(1+\varepsilon)+\beta\ell_2)=1$. If $u'(c)$ is eventually inelastic, then ℓ_2 must fall as β rises, and producer B will be forced out of the information good market. However, with good 2 as numeraire, it will always be the case that B will earn 1 and A will earn $1+\varepsilon$.

With an indivisibility, however, the situation is quite different. Suppose that it costs a fixed amount C to operate in the information good market at all. When ℓ_2 falls below C producer B no longer finds it profitable to participate in the information goods market and drops out entirely. This occurs when $\beta u'(\beta(1+\varepsilon)+\beta C)=1$. In this case producer B of course continues to earn 1. However, prices in the information goods market now jump to $\beta u'(\beta(1+\varepsilon))$, and producer A now earns $\beta u'(\beta(1+\varepsilon))(1+\varepsilon)$, which will be significantly larger than $1+\varepsilon$.

The argument can easily be generalized to a dynamic setting with capital accumulation, endogenous labor supply, and so forth. It shows quite starkly that, under very common circumstances, the simplest kind of technological progress may have a non-monotone and non-homogeneous impact on the wage rate of different kinds of labor. Our model predicts that continuing improvements in the technology for reproducing "information goods" have a non-monotone impact on wages and income inequality among producers of such goods. Initially, technological improvements are beneficial to everybody and the real wage increases at a uniform rate for all types of labor. Eventually, though, further improvements in the reproduction technology lead to a "crowding out" of the least efficient workers. When the process is taken to its natural limit, this kind of technological change has a disproportionate effect on the best workers. For large values of β, the superstar captures the whole market and has earnings that are no longer proportionate to the quality of the good it produces or its skill differentials, which are only slightly better than average.

To an external observer the transition between the two regimes may suggest a momentous change in one or more of the underlying fundamentals. In particular, one may be lead to conclude that the observed change in the dynamics of skill premia is due either to a shift from neutral to "skill biased" technological progress, or to a dramatic variation in the relative supply of the two kinds of labor, or, finally, to large changes in the skill differentials of the two groups. These are the main interpretations that a large body of recent literature has advanced to understand the evolution of wages during the last twenty-five years. While one or more of these explanations may well be relevant, our simple example shows it needs not be and, we would argue, it certainly is not for those sectors in which "information goods" are produced. We find the explanation outlined here not only simpler but also, plainly, more realistic.

Our point of view puts at the center stage the working of competitive forces when there is indivisibility and the unavoidable consequences of the law of comparative advantages. Our theory predicts that even very small skill differentials can be greatly magnified by the easiness with which information can be reproduced and distributed. It also predicts that the increased reproducibility of information will continue generating large income disparities among individuals of very similar skills and in a growing number of industries.

6. Debating Common Fallacies

Common legal, and often economic, wisdom argues that competitive markets are not suitable for trading copies of ideas, as ideas are intrinsically different from other economic commodities. For the most part these arguments are incorrect. Along most dimensions, ideas are not different from other commodities, and those few dimensions along which ideas are different do not generally affect the functioning of competitive markets. Here are some often-heard arguments, which we have shown to be fallacious.

It is argued that in competitive markets innovators would be unable to appropriate more than an infinitesimal share of the social value of their ideas. This is a recurrent theme in much business, managerial, and industrial organization literature, where it is apparently believed that economic efficiency requires innovators (or producers more generally, we would believe) to appropriate all the social value of their products. Were this to be the case, any market transaction in which some positive social surplus is realized would be inefficient as producers are "leaving something on the table", to consumers in fact. But, obviously, socially efficient provision of ideas/goods requires, instead, that all ideas/goods with a positive social surplus (i.e. social value larger or at most equal than social cost) be produced. How such surplus is split between producers, consumers, and other entities (suppliers of intermediate inputs, government, etcetera) may, and in general will, affect if all goods with positive social surplus are produced, but there is no general presumption that too few goods will be created unless producers appropriate the whole social surplus. In general, in fact, we would expect producers to bring goods, or ideas, to the market, as long as the private costs of doing so are exceeded by the private gains. Hence, from a social perspective, one should ask: for all ideas with a positive social surplus, is it the case that competitive pricing allows producers to appropriate enough revenues to compensate for their opportunity cost? Strangely enough, this question is seldom asked in the theoretical literature on innovations, and never, to the best of our knowledge, in the empirical one. This fallacy, as we have shown in Section 3, misses the fact that ideas combine attributes of both consumption and capital goods. They can be used directly for consumption, such as reading a book, or watching a movie, or they can be used as an input in production, by making copies of a book or movie, or by producing other goods, for example, by using the idea for an improved production process. That the original copy of an idea is the capital good (the tree) from which all other copies (the fruits) must originate enables innovators to appropriate the net present value of all future copies through competitive pricing. Corn seeds, for example, can be eaten or used for producing additional corn, so also combine characteristics of consumption and capital goods. Competitive markets for corn generate the appropriate incentive to invest in corn seed. The initial copy (or copies, when simultaneous innovation occurs) of an idea is generally produced through a process which is different from the one used to make subsequent copies, as in the case of original research versus teaching. Most capital goods (original research) are used to produce commodities other than themselves – but the fact that capital goods might be used to reproduce themselves poses no particular problem for competitive markets. In the semi-conductor industry, for example, reduction in chip size makes it possible to construct capital equipment that can be used to produce even smaller chips.

There are suggestions that ideas are subject to "spillover externalities", or what we might call informational leakage. That is, the existence of the idea enables people to learn it and make use of it without the permission of the owners. Some even argue that ideas can be copied for free. In practice, few ideas are subject to informational leakage, and in all cases are costly to reproduce. In the case of copyrightable creations, where the ideas are embodied in physical objects such as books, informational leakage is not an issue. In the case of scientific advances, reflection shows that it is also not the case. While in some sense scientific ideas are widely available, usable copies of scientific ideas are not so easy to come by. Even Newton's laws, our example in Section 2, require a substantial amount of time and effort to understand. For all practical purposes copies are limited to those people who understand the laws and books that explain them. Without paying someone to teach you or buying a book that explains Newton's laws, you are not terribly likely to learn them merely because they are in the public domain. As teachers and professors we earn our living by our ability to communicate ideas to others, and in doing so creating new copies of them. Overwhelming historical evidence shows that diffusion and adoption of innovations is costly and time consuming.

Leaving ideas in the public domain, as it would be the case under a NIP system, is socially inefficient and leads to a "tragedy of the commons" for creative activity. Although legal scholars have tended to view the public domain as a commons, like the atmosphere or ocean for which there are no property rights, in fact the market for a public domain book is very similar to the market for wheat or any other competitively provided good or service. Once copyright has expired, there are many copies of a book, each a good substitute for the other, and each owned by someone. If you want to read the book, make copies, or turn it into a movie, you must first buy the book from one of the current owners. If there are many owners, each competing with each other to sell you the book, you may be able to obtain it relatively cheaply, even though you intend to turn it into a highly valued movie. But the fact that you can buy ingredients cheaply is a good consequence of competitive markets, not a bad one. In fact, the evidence suggests that the market for goods in the public domain functions well, with copies widely available and reasonably priced: finding a copy of a book by Dickens, for example, is no great problem.

Lawyers have also made other arguments, based on the "public domain is like the commons" fallacy, as to why ideas might be different from other goods; but many of these arguments reflect lack of understanding of how markets function. For example, it is often argued that without the monopoly provided by copyright, there would be an inadequate incentive to "promote" works such as books, music and movies, since the benefit of the promotional effort would be shared by competitors. However, this argument applies equally well to other competitive markets, such as that for wheat. The point to understand is that under monopoly, goods are priced high, and the consumer receives little benefit. Hence the monopolist has an incentive to subsidize information to the consumer. In competitive markets, the competitors do not have incentive to subsidize information, so consumers must pay the cost of obtaining it. Information about wheat is widely available – from doctors, diet advisers, books, magazines, and many other sources – but not directly from wheat producers. In competitive markets, not only is information widely available, but it is less biased than the subsidized information provided by monopolists. Markets for ideas are no different in this respect. Plentiful information is available about works in the public domain – but that information is not generally provided by book publishers.

7. Conclusions

While the functioning of competition in the market for goods has been the subject of study for a long time, and our knowledge of the subject appears to have progressed substantially since the times of Adam Smith, it is often felt that the same is not true of the market for ideas. Indeed, there is a widespread view that ideas are dramatically and intrinsically different from goods and that the "economics of knowledge" needs to be grounded on different premises and adopt different modeling strategies than the rest of economics. In our work we reconsidered this issue and concluded that, while the economic theory of ideas does require modifications in some of the more common assumptions with which markets for regular commodities are handled, such differences are much less dramatic than one would have expected prima facie, and that a great deal of common economic wisdom applies equally well to the economics of knowledge. This allows us to critically reconsider a number of theoretical issues sitting at the intersection between the theory of innovation and technological change and growth and trade theory, to conclude that much "new common wisdom" (otherwise known as "new growth theory" or, in its extreme version, "new economy") is either empirically groundless or logically faulty, and that some old, and possibly uncommon, wisdom, should be brought back to bear on the study of technological change, growth, and trade. Central to understanding the market for ideas and the incentives for the adoption of new ideas is discovering how ideas might be different from other goods. The starting point of the economic analysis of innovation is to recognize that the economically relevant unit is a copy of an idea. That is, typically, many copies of an idea exist in physical form, such as a book, a computer file or a piece of equipment, or in the form of knowledge embodied in people who know and understand the idea. When embodied in humans, copies of ideas are labeled with a variety of different names, which often obscure their common nature: skills, knowledge, human capital, norms, and so on. Careful inspection shows, though, that each and everyone of these apparently different entities is, at the end, nothing but the embodied copy of an idea, and that the latter was either discovered first by the person in whom it is currently embodied, or costly acquired (possibly via observation and imitation) from other humans, in whom it had been previously and similarly embodied. Economically valuable copies of ideas do not fall from the heavens, like manna, but are the product of intentional and costly human efforts. Only these copies matter, first, in the sense that if they were all to be erased, the idea would no longer have any economic value, and, second, in the sense that the copies are relatively good substitutes for each other: whether a copy of an idea is the original copy or the hundredth copy, it is equally economically useful. From the perspective of the functioning of markets, then, property rights in copies of ideas is assured by the ordinary laws against theft – what is ordinarily referred to as "intellectual property" protects not the ownership of copies of ideas, but rather a monopoly over how other people make use of their copies of an idea.

References

Arrow, K.J. (1962), "Economic Welfare and the Allocation of Resources for Invention", in Richard Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton NJ, Princeton University Press.

Besen, S. M., and Kirby, S. N. (1989), "Private Copying, Appropriability, and Optimal Copying Royalties", *Journal of Law and Economics*, 32; 255-80.

Boldrin, M. and D. K. Levine (1999), "Perfectly Competitive Innovation", PDF version available at www.dlevine.com and www.econ.umn.edu/˜mboldrin.

Boldrin, M. and D. K. Levine (2002), "The Case Against Intellectual Property", *American Economic Review*, Papers and Proceedings, 92; 209-12.

Boldrin, M. and D. K. Levine (2004a), "Rent Seeking and Innovation", *Journal of Monetary Economics*, 51; 127-60.

Boldrin, M. and D. K. Levine (2004b), "IER Lawrence Klein Lecture: The Case Against Intellectual Monopoly", *International Economic Review*, 45; 327-50.

Boldrin, M. and D. K. Levine (2004c), "IP and Market Size", PDF version available at www.dlevine.com and www.econ.umn.edu/˜mboldrin.

Boldrin, M. and D.K. Levine (2005a), "The Economics of Ideas and Intellectual Property", *Proceedings of the National Academy of Sciences*, 102; 1252-56.

Boldrin, M. and D.K. Levine (2005b), "The True Social Cost of Intellectual Monopoly", in process, University of Minnesota and UCLA.

Caves, R.E., M.D. Whinston and M.A. Hurwitz (1991), "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry", *Brookings Papers on Economic Activity*, Microeconomics; 1-48.

Congressional Budget Office (1998), "How Increased Competition from Generic Drugs has Affected Prices and Returns in the Pharmaceutical Industry", The Congress of United States, mimeo.

DiMasi, J.A., R.W. Hansen, H.G. Grabowski and L. Lasagna (1991), "The Cost of Innovation in the Pharmaceutical Industry", *Journal of Health Economics*, 10; 107-42.

Johnson,W. R. (1985), "The Economics of Copying" *Journal of Political Economy*, 93; 158-74.

Klein, B., Lerner, A. V., and Murphy, K. M. (2002), "The Economics of Copyright: 'Fair Use' in a Networked World", *American Economic Review*, Papers and Proceedings, 92; 205-08.

Legros, P. (2005), "Art and Internet: Blessing the Curse?", mimeo, ECARES, Université Libre de Bruxelles and CEPR, February.

Liebowitz, S. J. (1985), "Copying and Indirect Appropriability: Photocopying of Journals", *Journal of Political Economy*, 93; 945-57.

Nordhaus, W. D. (1969), *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change.* MIT Press, Cambridge.

Ordover, J. A., and Willig, R. D. (1978), "On the Optimal Provision of Journals qua Sometimes Shared Goods", *American Economic Review*, 68; 324-38.

Plant, A. (1934), "The Economic Aspects of Copyright in Books" , *Economica*, 1; 167-195.

Romer, P.M. (1986), "Increasing Returns and Long Run Growth" , *Journal of Political Economy*, 94; 1002-37.

Rosen, S. (1981), "The Economics of Superstars", *American Economic Review*, 71; 845-58.

Scotchmer, S. (1991), "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law", *Journal of Economic Perspectives*, 5; 29-41.

Stigler, G. J. (1956), "Industrial Organization and Economic Practice," in L.D. White (ed.), *The State of the Social Sciences*, Chicago, University of Chicago Press.

Michele Boldrin (corresponding author); University of Minnesota, Federal Reserve Bank of Minneapolis, and CEPR London; mboldrin@econ.umn.edu. David K. Levine; UCLA and Federal Reserve Bank of Minneapolis, david@dklevine.com. Both authors thank the NSF for financial support.,