

The economics of ideas and intellectual property

Michele Boldrin^{†*} and David K. Levine[§]

[†]Department of Economics, University of Minnesota, Minneapolis, MN 55455; and [§]Department of Economics, University of California, Los Angeles, CA 90095

Edited by Thomas J. Sargent, New York University, New York, NY, and approved December 10, 2004 (received for review October 18, 2004)

Innovation and the adoption of new ideas is fundamental to economic progress. Here we examine the underlying economics of the market for ideas. From a positive perspective, we examine how such markets function with and without government intervention. From a normative perspective, we examine the pitfalls of existing institutions, and how they might be improved. We highlight recent research by us and others challenging the notion that government awards of monopoly through patents and copyright are “the way” to provide appropriate incentives for innovation.

economic theory | game theory

Central to understanding the market for ideas and the incentives for the adoption of new ideas is understanding 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. 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 extremely 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.

Common legal and 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. However, there is one dimension in which economists agree that ideas differ in an important way from other goods: The first copy of an idea must generally be produced as a single indivisible unit. For example, two first halves of a book are not a good substitute for both the first and second half. Although most goods are subject to some degree of indivisibility, quantitatively, the indivisibility is crucial for ideas. In much of the economics literature, especially in Romer (1), the cost of this indivisibility is referred to as a fixed cost to emphasize the fact that it is paid once regardless of how many copies are later produced. The key economic question this raises is whether in competitive markets there is adequate incentive to produce the first copy of an idea. Earlier work, such as that of Romer (1), suggests that because after the fixed cost is incurred without legal protection sales take place at marginal cost, leaving no profit to recoup the fixed cost, the answer to the question is negative; without legal protection, there is inadequate incentive to innovate. However, we point out that this is the case only if one assumes that either marginal costs are zero (copies of ideas are perfectly nonrivalrous goods) or capacity constraints are not binding even immediately after the innovation takes place. In reality, and in the theory we develop, copies of ideas are rivalrous goods and there are generally capacity constraints; hence, unless the innovator and his competitors can instantly flood the market with copies, ideas will always generate

some rents for their creator. This gives rise to a different theory, where innovations may or may not arise in competitive markets depending on the size of the initial indivisibility or fixed cost, the speed at which capacity is built up, the size of the market, and the elasticity of demand. From a practical perspective, the issue is whether competitive rents are enough to give creators the incentive to invent an adequate number of socially valuable ideas. Making creators fabulously rich is not a relevant issue, unless this is an essential part of the incentive to create. Compensating creators for the opportunity cost creative activity involves is the economic issue to be addressed by the normative theory of innovation.

Along most other 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.

1. 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 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.
2. The initial copy (or copies, when simultaneous innovation occurs) of an idea are generally produced through a process that 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 semiconductor industry, for example, reduction in chip size makes it possible to construct capital equipment that can be used to produce even smaller chips.
3. 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 original inventor. Some even argue that ideas can be copied for free. In practice, few ideas are subject to informational leakage, and in all cases they 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

This paper was submitted directly (Track II) to the PNAS office.

^{†*}To whom correspondence should be addressed. E-mail: mboldrin@econ.umn.edu.

© 2005 by The National Academy of Sciences of the USA

to be solved that is not already solved by the competitive marketplace.

The problem that arises, as we have indicated, is that of indivisibility. Assume this is such that a choice of $x_0 < 1$ implies no innovation at all. It may be, depending on conditions of demand relative to μ , that the optimal initial choice of capital for a competitive innovator is $x_0 < 1$, so that this indivisibility binds. In this case, the only realistic option is to choose $x_0 = 1$. Although the creator still receives a positive revenue of P , it is insufficient to compensate for the cost of creation μ and so the good is not produced. We should indicate that although it is easy to work out the consequences of a binding indivisibility in this simple setting, general equilibrium theorists have yet to create a comprehensive theory of competitive equilibrium with binding indivisibilities.

In the case of indivisibility, the issue is whether the revenue stream, P , evaluated at the minimum innovation size, $x_0 = 1$, is sufficient to compensate for the cost of creation, μ . In the extreme case in which $\beta \rightarrow \infty$, revenue $P \rightarrow 0$; in this case, no innovation would take place at all. It is a not-uncommon confusion to believe that this limit case is in fact the ordinary case. Both theory and evidence suggest that competition generates substantial revenues for innovators in most practical cases and that binding indivisibilities are the exception, not the rule. Note also a second common source of confusion: The revenue stream, P , accruing to the innovator almost never corresponds to the full social value of the new idea. This is more so when the invisibility $x_0 \geq 1$ is binding, and P may be a relatively small fraction of the total additional utility the innovation will bring to society. But this fact is of no concern for economic efficiency, as long as $P \geq \mu$ holds; institutions that allow creators to be compensated for the opportunity cost of their effort yield socially efficient outcomes. Making creators extremely wealthy is a by-product that is welcome to them but unnecessary to society.

Remark. The creator may have a unique idea, in which case he faces no competition for providing it. He is a monopolist in the initial period and maximizes the objective function, $\lambda \int_0^\infty e^{-u'}(xe^{\beta t})xe^{\beta t} dt - \mu x$, with respect to x . Notice that here he no longer takes present and future prices, $u'(xe^{\beta t})$, as given, and the solution to this problem is generally to produce too few initial copies of the idea. However, in the case where the indivisibility binds, the producer is still forced to provide a single initial copy, and his monopoly over the unique idea is irrelevant.

Results

First-Mover Advantages. 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 x_t to expand so rapidly as to flood the market and drive price to zero almost immediately. As we observed, as $\beta \rightarrow \infty$, the revenue $P \rightarrow 0$. It is worth noting that the same technological change is reducing the cost of books, music, and movies creation as well, so that $\mu \rightarrow 0$ also, 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 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 3 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. Evidence from the pharmaceu-

tical 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 >50% of the market (6, 7). This is, of course, not decisive evidence: It could be that the monopoly provided by the patent is important in building consumer loyalty that persists after the patent expires. In the case of financial securities, we do have stronger evidence on the presence of a first-mover advantage; until the State Street Decision in 1998 financial securities could not be patented, and as documented by Tofuno (8) among others, there was thriving innovation driven by a strong first-mover advantage. Although imitation was rapid, the first mover successfully maintained the bulk of the market against imitators.

The pharmaceutical industry makes much more extensive use of patents than other industries, and the expense of bringing a new drug to market, including the cost of clinical trials and failures, is estimated by DiMasi *et al.* (9) at \$231 million 1987 dollars. Would not, as the industry argues, eliminating patents in that industry cause innovation to come to a screeching halt? We should point out first that patents are only one part of government regulation of pharmaceuticals, the FDA supervision of clinical trials, the subsidy of basic research in the area, and the large government purchases of drugs being other key elements. It would be hard to make sense of a proposal that would eliminate patent protection while allowing competitors to freely make use of the results of expensive clinical trials. However, if competitors were required to choose between either purchasing (possibly at a price regulated by the government) the results of trials or conducting their own trials, a substantial first-mover advantage would be preserved. The cost of research prior to trials is heavily subsidized, and the subsidy could be increased. It should also be recognized, as we discuss below, that the elimination of patents would have a positive effect on innovation, by allowing researchers to freely use the results of others without the need to obtain patent clearances. One almost certain effect is that it would eliminate the considerable effort wasted in inventing “work-alike” drugs in an effort to share in a lucrative monopoly.

The first-mover advantage is a form of monopoly accruing to the original innovator. A monopolist, unlike a competitor, will not allow quantity x_t to expand to x^c , which drops the price to zero, but will restrict output to $x^c/2$, which maximizes his revenues. In the limit, as $\beta \rightarrow \infty$, output jumps almost immediately to $x^c/2$, resulting in a revenue to the monopolist of $\lambda\rho$. If this revenue can be captured only for a fraction of time, ϕ , then the corresponding revenue is $\phi\lambda\rho$. For computational simplicity, we will focus hereafter on the case of β large, although it considerably understates the benefits of competition. The first-mover advantage here may be represented by a fraction, ϕ^F , representing time before competitors are able to successfully enter. Patent and copyright monopolies can be represented by a fraction, $\phi > \phi^F$, representing the duration of the legal protection.

We have seen how, under some circumstances, there may be underprovision of ideas due to indivisibility. We turn now to the traditional solution to this problem: 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 greater incentive to create or innovate. There are, however, a number of problems with this solution.

Sequential Innovation. Because a monopolist is scarcely likely to earn less than a competitor, it might seem that whatever the

problems associated with monopoly, government grants of monopoly for innovation at least increase the incentive to innovate. But, just as most commodities are produced by means of other commodities, so are ideas; innovations build on past innovations—so while raising the profit from innovation, granting monopoly on newly created ideas also raises the cost of future new ideas.

Sequential innovation and the way in which patents inhibit innovation have been studied by Scotchmer (10) and by Boldrin and Levine (3, 5). We illustrate this with a simple example, collapsing the dynamic model introduced above into a static one in which $\beta = \infty$. Utility continues to be $2\rho[2(x/x^c) - (x/x^c)^2]$. A monopolist will produce $x^c/2$, resulting in a revenue of $\lambda\rho$. A competitive innovator with a first mover advantage will get a revenue of $\phi^F\lambda\rho$. Suppose that producing the new idea requires the use of N existing ideas. We imagine that each of these many ideas is small, so that the cost of producing a copy of the idea is ε/N . Without government monopoly, 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 $\phi^F\lambda\rho > \varepsilon$.

Suppose, on the other hand, that the government-awarded monopoly applies to all innovations, and that the owners of the N existing ideas only know that ρ is drawn from a uniform distribution over $[0, \bar{\rho}]$. Each sets a price, p_i , at which they will license their 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

$$\frac{\lambda\bar{\rho} - (N - 1)p - p_i}{\lambda\bar{\rho}} p_i.$$

If $\varepsilon < \lambda\bar{\rho}/2$, the Nash equilibrium of this game is at $p = \lambda\bar{\rho}/(N + 1)$, and therefore the inventor must pay $(N/N + 1)\lambda\bar{\rho}$ to clear the needed rights for his own innovation, and so he innovates if he draws an innovation for which

$$\rho > \frac{N}{N + 1}\bar{\rho}.$$

This occurs with probability $1/(N + 1)$. By way of contrast, without monopoly the probability of innovation is $1 - \varepsilon/(\phi^F\lambda\bar{\rho})$. As the number of existing rights that must be cleared increases, the probability of innovation under monopoly is smaller than that under competition and drops towards zero. Here 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.

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 bestseller 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 United States, 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 whereas 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 they replace trade-secrecy and force 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. And because patents last 20 years, the only reason to get a patent is that the inventor thinks he cannot keep the secret for that long. We have studied this issue in ref. 11, showing that creating public rent-seeking is not a good way to solve the problem of private rent-seeking.

Optimal Duration of Intellectual Monopoly. Although intellectual monopoly may encourage socially desirable innovation, it has a number of drawbacks, as we have seen. These range from the traditional fact that monopolies overprice and undersupply to rent-seeking and the discouragement of subsequent innovation. To understand more clearly the trade-off involved with government awards of monopoly, we examine a simple example in which we abstract from rent-seeking, sequential innovation, and competitive rents. We focus only on the traditional monopoly undersupply, and ask: In a world in which ideas are of variable quality, what is the optimal level of protection, ϕ ? Unlike an earlier economics literature pioneered by Gilbert and Shapiro (12), and discussed at length in Gallini and Scotchmer (13), we follow Grossman and Lai (14) and Boldrin and Levine (15) in taking a general equilibrium approach in which there are many innovations.

Suppose the first copy of any idea has a unit cost of creation, and the factor used in producing the first copy is abundant and inelastically supplied; reproduction costs are zero. We continue to use $u(x) = 2\rho[2(x/x^c) - (x/x^c)^2]$ for consumer utility. The social value of the idea under monopoly is $(3/2)\rho$ and under competition 2ρ . Under monopoly, revenue equals ρ ; hence, the latter is also a measure of the private value of a good for the monopolistic innovator. We let ϕ denote the fraction of the time the producer has a monopoly; without government intervention this is ϕ^F due to first-mover advantage. By providing copyright and/or patent protection, the government can raise ϕ to any higher value up to a limit of 1. There are λ consumers. Then ideas will be produced for which private revenue exceeds cost; that is, $\phi\lambda\rho \geq 1$. In particular, without government intervention, so $\phi = \phi^F$, as the size of the economy λ grows, the quality of the marginal idea that is produced, $\rho = 1/\phi^F\lambda$, declines and more ideas are produced. We continue to suppose that ideas are uniformly distributed on $[0, \bar{\rho}]$. We must set $\phi > 1/(\lambda\bar{\rho})$ if any ideas are to be produced at all. Assuming this is the case, social welfare will be

