Chapter 2: The Theory of Innovation without Intellectual Monopoly

The traditional perspective of economists is that useful ideas lead a disembodied existence. Ideas are not rivalrous – once discovered, they are akin to a public good. As a result, in a competitive market, ideas trade at a zero price, the entrepreneur cannot recoup the fixed cost of innovating and intellectual monopoly is seen as a necessary evil. Our perspective is that ideas are useful only to the extent that they are embodied in people or things, and that as a result once discovered they are not a public good and must trade at a positive price. Although we do not view the cost of producing an idea as a fixed cost, it does involve an indivisibility, and under certain circumstances this does lead to the traditional conclusion that a particular idea may not be produced in the absence of intellectual monopoly. Our immediate goal is to understand how prices of ideas are determined in the absence of intellectual monopoly, and consequently which ideas will and will not be produced in its absence.

An Example

The issue of whether ideas are rivalrous or non-rivalrous; embodied or disembodied; whether costs are fixed or whether there is indivisibility are all rather abstract. To understand the practical significance of these different ways of thinking, we begin with an illustrative example.

Think about the problem of developing a new drug. Perhaps a team of twelve expert biomedical researchers working full-time for one year will be able to create the new drug. Now some would argue that the significance of the new drug is a formula written on a sheet of paper that can costlessly be transmitted to, and understood by anyone. This is the traditional notion that ideas are *disembodied* – that they lead an independent existence. Another way to say this is to say that ideas are *non-rivalrous* or a *public good* – meaning that one person reading the sheet of paper and using the idea does not prevent someone else from doing the same thing at the same time. We would prefer to emphasize that the process of research generates a great deal of useful knowledge beyond a simple chemical formula – practical knowledge about dosages and about methods of manufacture. This is not simple abstract knowledge with an independent existence, but exists only in the minds of the research team – it is *embodied* in that team: In order to build a facility for mass production, the team is going to have to explain to lots of other people how to manufacture the drug. Simply reading the formula, even assuming that such formulas were easy to read and understand, would not do.

Implicit in the view that ideas are embodied is the fact that the communication of ideas is costly. Even if the only relevant knowledge is the formula on the sheet of paper, if the team has to explain it to you or me it is going to take quite a while, since first we will have to get twelve different doctorates so we can understand what the formula means. For other biomedical research teams, the time to explain may be rather faster. In the case in which the knowledge is not a mere formula on a scrap of paper, but technical knowledge about how a particular disease mechanism works, and details about how a drug can be manufactured, perhaps what took the original team a year to develop could be explained through a month or even a week of teaching to a second team of experts. In either case: it will take valuable time and resources until a second team is capable of manufacturing the drug.

The critical difference between the economic implications of this point of view and of the conventional one is whether ideas will sell for a positive price in the absence of intellectual monopoly. According to the disembodied/non-rivalrous/public good perspective, ideas must sell for a zero price because once put into use they are available to everyone for free. We think this is naïve. Ideas must always sell for a positive price, because they are expensive to communicate, and because there is a social value in spreading ideas. In the case of drug production, it may be valuable to have more than one group of 12 people who know how to produce the new drug. With this we mean that there is positive social value in the second research team because, for example, production can take place in parallel, or one team can set up a facility in the US, while the other sets up one in Europe. In competitive equilibrium the price of an idea is equal to the additional social value from producing more of it. Since transmitting the idea to a second team has positive social value, the implication is that, even in the absence of intellectual monopoly, the idea sells for a positive price.

The general principle we are trying to convey is that ideas always have a positive price because they are costly to transmit, learn and implement. Economists have somehow failed to gather the relevance of this point when thinking about innovation and economic growth, while they have stressed it repeatedly when thinking about private information, optimal contracts, financial markets, technology transfer and so on. Such asymmetry is surprising; its endurance in the face of overwhelming evidence to the contrary, even more so. Its roots are to be found, most probably, in a metaphor popular among general equilibrium theorists and macroeconomists. According to this metaphor, economically valuable ideas are like recipes: they contain exhaustive instructions on how to do things. Economists, clearly, must not be familiar with cooking. Only someone who seldom sets foot in a kitchen may think that all that there is to good cuisine is a very detailed recipe.

Still the question remains whether the price of an idea is sufficient to cover the cost of producing it. In the traditional view, the cost of producing an idea is viewed as a *fixed cost*, meaning that no production can take place until the fixed cost is paid. We would prefer to emphasize that producing an idea is similar to producing any other capital good. After the idea is produced, we now have – an idea. The person or thing in which this idea is now embodied can, like any other capital good, produce other ideas, or be used as a template to produce goods and services, in this instance, drugs. However, in contrast to other capital goods that can often be produced in variable quantities, an idea must generally be produced whole. Two first halves of an idea for a drug are certainly not worth the same as the first and second half of the idea. We describe this by saying that there is an *indivisibility* in the production of ideas.

This may seem like a small distinction – there is agreement in both points of view that there is an indivisible cost (one year of team time) in producing the research, and that as a result there is now the possibility of producing a drug of economic value. But when the one year of team time is seen as a fixed cost, it implies that nothing able to fetch a positive price in the market has been obtained after the year has elapsed. In the fixed cost view a successful research effort simply generates a nonrivalrous idea which, without patent protection, can be grabbed by anyone. Hence competitive pricing cannot do the job, almost by construction: after one year of search we have nothing saleable in our hands. However, and unlike a fixed cost, an indivisibility might or might not make a difference for competitive pricing. First of all, the indivisibility point of view says that when the research effort is successful we have something in our hand, which is saleable: the template of a new idea. Problems may arise for competitive pricing because even this first indivisible template may be too big for the market to bear. That it "may be too big" does not imply, though, that it always is. In fact, most often it is not. For example, shoes must be produced in whole units, and as a result we may have only 1,000,000 pairs produced when in fact it is socially desirable to have 1,000,001 pairs produced – but quite rightly no one worries about this. Similarly, the minimum seat capacity at which a commercial jet can be manufactured may be of 20 seats, but again this may not matter much as most airlines demand commercial jets with 20 seats or more. In our simple example, once the team time has been paid for we have a marketable object: a team that can either produce the drug or teach other teams how to produce the drug. The first team fetches a positive competitive price because having a second team still has private and social value. Focusing on the fixed cost rather than the indivisibility tends to obscure the fact that the indivisibility might not matter much in most practical circumstances.

Nevertheless, sometime the indivisibility does matter, and we should look carefully into the cases in which it does. Social *efficiency* requires that whatever has a social value greater than or equal to its social cost, be produced. The indivisibility is a technological restriction, without which it might be socially desirable to have only half a drug produced. Pretend the idea of a drug was perfectly divisible, so even a fraction of it had positive value. For society the drug might be worth $8B when it costs half a team $401M to discover half of it. Having an entire team knowing about the drug might increase social value to $8.1B, while costing an additional $401M. So, of course, it would not be worth paying $401M for a gain of only $100M and private markets would, quite properly, produce only the first half of the drug. With divisibility of drug ideas competitive markets can do the job: discover and produce the half drug. However, given the indivisibility, the social choice is between either spending nothing, and staying without drug, or paying $802M for a drug worth $8.1B – clearly a socially worthwhile investment, given that the half option is not available. Unfortunately, competitive markets do not find this social point of view all that compelling. In this case the indivisibility binds for competitive markets. Let us see in further details why this is so.

Social efficiency requires comparing total social value to total social cost, while the price at which a good sells under competition is its *marginal private value*, and the latter is compared by the private entrepreneur to the *marginal private cost*. The former may be smaller than the latter when going from, say, almost-one-idea to one-full-idea; in the example above the additional cost of going from half to a full idea is $401M while the gain is just $100M: competitive markets would not produce the drug, and this is socially inefficient. The larger is the size of the indivisibility, relative to the size of the economy, the more likely this outcome is. We concur, in other words, with the traditional wisdom that there may be socially valuable ideas that will not be produced under competition. We will consider later whether, under the same circumstances, they will actually be produced under monopoly and which remedies, if any, should be undertaken.

On the other hand, and in contrast to the conventional wisdom, we recognize also that there are cases where ideas will be produced under competition. If the social optimum is to have more than one team doing research in parallel, for example, then it must be that the social benefit of the second team exceeds its cost, meaning that the first team can sell the idea for a price that covers its own cost. In this case, competitive markets are able to reach the socially efficient outcome while monopoly would hinder it.

Now it may be puzzling to think of having two teams researching in parallel. Would it not be better to have the first team do the research, then teach the second team? This would use one year and two months of team time to have two fully trained teams, rather than two years of team time. However, if the two teams research in parallel, the drug can be produced (on a double scale) after one year, rather than after one year one month. So the question is, does the social benefit of having the extra amount of drug available one month earlier exceed the social cost of ten months of team time? If so, at the social optimum, the drug would be simultaneously engineered and, after one year of calendar time, production would start at twice the capacity. In this case at least one of the two teams can cover its costs under competition.

Notice a second and very important consequence of this point of view, also not easily seen in the fixed cost point of view. As the economy expands in size, the economic relevance of the indivisibility is progressively reduced. Moreover, as there are more and richer people to benefit, the value of early availability increases. And so economic progress diminishes the rationale for intellectual monopoly.

The Conventional Wisdom

That the downstream licensing provisions of patent, copyright and other private contracts leads to monopoly is well understood by economists. The argument, as we have seen, is that only through monopoly is it possible to reward inventive activity. We reiterate the traditional logic. The cost of innovation is a fixed cost. Once discovered, ideas are a public good reproduced and distributed at constant marginal cost. Since perfect competition prices at marginal cost, profits are zero, and the fixed cost cannot be recouped. Consequently, without intellectual monopoly, there will be no innovation. This idea forms the foundation for a wide variety of economic models, ranging from general equilibrium models of monopolistic competitions such as Romer, and Grossman and Helpman, or Aghion and Howitt, to micro-models of patents and patent races such as Gilbert and Shapiro, or Gallini and Scotchmer. The original theoretical argument was sketched by Alwyn Young before the Second World War and developed in greater detail by Joseph Schumpeter right after the war. The first formal treatment of the idea that competitive markets are intrinsically incapable of handling innovations can be found in writings by Arrow and subsequently Shell, published in the early and middle sixties.

To provide microeconomic foundations to the notion that, once discovered, ideas are akin to public goods Romer introduces the distinction between *rival* and *non-rival* goods. Rival goods are like potatoes: if I consume a potato, you cannot consume the same potato. Non-rival goods are like ideas – my enjoyment of the fundamental theorem of calculus does not in any way interfere with your enjoyment of the same theorem. Economists further distinguish between *excludable* and *non-excludable* goods. Ideas are an example of an excludable good: the mere existence of an idea does not mean that everyone must be allowed to make use of it. A standard example of a non-excludable good is the military: the military cannot easily defend my house, without also defending yours.

Public goods are generally defined to be non-exclusionary and non-rival, such as the military: economists have long argued that only government can provide such goods, either directly or through subsidies. Non-rival goods that are excludable are thought to present less of a problem: they can be privately provided, with profit generated through the exclusion of low valued users. Under informational constraints about who the low valued users are, this may be the best obtainable result: in economic parlance the *second best*. Many municipal utilities are thought to fall into the non-rival excludable class, including electricity, water, telephone, cable, and such services as garbage collection. We are not going to debate if such goods are truly nonrivalrous; for the most part, we would argue, they are rivalrous. It suffices to note that, in practice, such services are governmentally provided in some municipalities, and privately provided in others. However: in those cases where services are provided by private monopolists, we find also that these monopolists are publicly regulated, generally by a utilities commission that has the power to set prices. Intellectual monopoly is relatively unique, in that the government grants a monopoly, but, with the exception of a handful of mandatory licensing statutes, does not regulate price or quantity.

The conventional argument for the necessity of intellectual monopoly does not depend on the fact that ideas are distributed at zero marginal cost; and indeed in the case of municipal utilities, the marginal cost of distribution is clearly positive. The problem is that with a fixed cost and constant marginal cost, there is overall increasing returns to scale, meaning that the total cost of each additional unit is less than that of the preceding unit. Since competition leads to a price equal to the marginal cost of the last unit produced, revenues would always be less than the total cost of production. With competition there is no way of charging for the original fixed cost as only the marginal cost is reflected in prices.

This argument is so powerful that it ought to in fact apply to all industries. For example, a factory that produces shoes requires a fixed cost to build. Once the factory is built, shoes can be produced at constant marginal cost, and so in a competitive industry of shoe manufacturers, no producer of shoes can earn enough to pay for the factory. Taking the argument to the extreme, we should then argue that the "solution" to this "problem" is for the government to give a monopoly to one shoe producer.

There is in fact no economic literature arguing that we must have government grants of monopoly power to one shoe manufacturer in order to have shoes. And the problem with the fixed cost/constant marginal cost argument is not difficult to see in the case of a shoe factory. A factory cannot produce an unlimited number of shoes. Rather, it can only produce shoes at constant marginal cost up to the capacity of the factory. If the factory is small enough, relative to the size of the market for shoes, it will produce only a modest number of shoes, and consumer will be willing to pay a premium over marginal cost for the limited number of shoes available.

This counter-argument against a shoe monopoly applies equally well to intellectual monopoly. Although the view is widespread that once discovered, ideas can be grabbed for free by anybody, this is far from the truth. While it may occasionally be true that an idea is acquired at no cost – I walk by a car, observe the wheels, and the idea of using wheels for transportation pops into my mind – ideas are generally difficult to communicate, and the resources for doing so are limited. It is rather ironic that a group of economists who are also college professors and earn a substantial living through the communication of ideas would argue otherwise. Take, for example, the famous idea $E = mc^2$. This is commonly known, in the sense that many people can quote the formula. But how many people actually know what it means, or can put it to any productive use? We will not belabor this point – once it is made it should be fairly self evident, and there is in fact an extensive literature (in economics, no less) on the problem of technology diffusion, that documents in substantial detail the costs of transferring ideas. In any case, we will be reviewing the evidence about the difficulty of transmitting ideas in a subsequent chapter.

This is the first junction at which our view departs from the conventional theory. As we have said, we believe that ideas have economic value only insofar as they are embodied in people or things – ideas do not lead an independent existence. Consequently there is always a cost of transferring an idea from one person to another, or of creating a duplicate object. Resources for communication and copying are always scarce. Ideas cannot be freely reproduced at an unlimited rate. In any given period of time, there are only a limited number of people who can get copies of an idea. As result, ideas are always scarce, and so they must sell for a positive price – even in the absence of intellectual monopoly. Because of this limited capacity, ideas will always sell above marginal cost. The difference between market price and marginal cost is referred to as a *competitive rent*. It accrues to the producer and, we will argue, it is likely to be enough to cover the initial cost of discovery. The link between competitive rents and limited capacity goes back to the work of Marshall, so this is not a new idea.

The fact that the producer will earn a rent means that some ideas may be produced, even in the absence of intellectual monopoly. But it does not imply that every socially valuable idea will be produced. Consider again the case of a shoe factory. The standard theory of competition, not only asserts that shoe factories will be built, but that the socially desirable number of shoe factories will be built. The reason for this is that shoe factories are divisible: we may build smaller or larger shoe factories. The builder of the factory, when deciding how large a factory to build, will not build so large a factory that the rents from the fixed capacity of the factory will be less than the cost of building the factory. The builder (facing competition from other shoe factory builders) will wish to increase the size of the factory as long as the rents from a little more capacity exceed the cost of adding the capacity. This is exactly the condition for a social optimum, and that is why economists do not argue that owners of shoe factories should be awarded government monopolies.

As we argued above with the example of a new drug, this does not necessarily apply in the case of intellectual property. Most ideas are not divisible. Two identical small shoe factories are a pretty good substitute for one large one. Two first halves of a book, are not a terribly good substitute for both halves. This indivisibility is the second key feature of the market for ideas.

The combination of embodiment of ideas and indivisibility does not mean that absence of intellectual monopoly is guaranteed to produce the socially optimal number of ideas. As we observed above, though, it does lead to rather different conclusions than the conventional view. To understand whether competitive rents are sufficient to cover the production cost of ideas we must understand the determinants of competitive rent. That is our next objective.

How Large Are Rents Without Intellectual Monopoly?

Much of the conventional confusion over intellectual property arises from an attempt to view a *dynamic process*, the invention, adoption and diffusion of a new good, through the lens of a *static model*, that of production and pricing in the presence of a fixed cost. In the model of fixed cost plus constant marginal cost unlimited capacity is taken for granted, and everything takes place at an instant of time: the new good is adopted and a large amount of it produced during one single period. When time enters these models (as for example in the modern work of Grossman and Helpman, Romer, and Aghion and Howitt among other) it is only because a whole sequence of goods is being invented, one in each period. After each good is invented, unlimited capacity is tacitly assumed for production of that good. In other words, conventional wisdom identifies the dynamics with the fact that new goods are invented, while interpreting the invention, adoption, and reproduction process for each single good as a static business. In reality, invention, adoption, and reproduction take place at different moments of time. The internal dynamics of each single episode of innovation is crucial to understand the economic forces behind it.

In particular: the good will be produced and sold only after it has been invented and after the human and physical capital needed to produce it have been installed and accumulated. For this reason we prefer to call the original cost of developing an idea a *sunk cost* rather than a *fixed cost,* and stress the fact that productive capacity is always limited. Consequently, while it is possible to eventually communicate an idea to an unlimited number of people, it is certainly not possible to reach more than a modest audience at any moment of time. Our drug team example also makes clear that only a limited number of people will be able to learn how to imitate an innovator during any particular period of time. Consequently, producers of ideas will earn a rent. How large this rent is, depends on how rapidly the new idea can be acquired by others, how quickly productive capacity can be installed, and how much value is derived from using the new idea. In contrast to earlier scientific writing in economics, the modern economics literature on innovation, such as our own work, and that of Hellwig and Irmen, takes the dynamic determination of rents as the point of departure.

A Mathematical Example

Assumptions

We use a simple example to calculate a lower bound on how much this competitive rent might be. More elaborated, and possibly more realistic cases, are introduced later in this and the following chapters.

Suppose that an invention has already taken place. In other words, there is currently either a single template item, book, song, or blueprint that is owned by the innovator, or a single team of employees of the innovator who have learned how to routinely produce the item. We focus first 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. We also restrict ourselves to the case in which the item requires only a copy of itself and no other inputs to be reproduced. The case in which the initial prototype and the actual consumption good are not perfect substitutes, and that in which inputs other than the good itself are required for reproduction, will be developed later.

Notice first that, even without intellectual monopoly, to the extent that it is difficult for purchasers to duplicate the object, the innovator will be able to exercise monopoly power, and earn a premium over the competitive rent. More generally, when the costs of imitation and reverse-engineering are substantial, the competitive rents we compute next represent only a lower bound on the revenues accruing to the innovator. We will examine these issues further in subsequent chapters.

At a moment in time, any resource has two alternative uses: it may be consumed or it may be used to produce more for the future. For almost all items these are mutually exclusive uses or, at the very least, consuming it reduces our ability of using a good for production during the same period. Nevertheless, we consider here the stylized and unrealistic case in which consumption and reproduction are mutually compatible at zero cost, and the traditional tradeoff between consumption today and consumption tomorrow disappears. As it will be clear momentarily, this is the worst possible scenario for our position as we allow for instantaneous and costless reverse engineering, and rule out the tradeoff between consumption and savings. Also, while the process of copying is resource consuming, we follow the existing economics literature in assuming that there is no other cost of producing copies. Without intellectual monopoly, as soon as the first unit is sold, the innovator is in competition with his customers. This means that we can derive the rent to the innovator through the ordinary theory of perfectly competitive equilibrium.

Details

Specifically, suppose that there are currently $k > 0$ units of the product available and that, independently of how many of these units are consumed during the current period, βk units of the same product will be available tomorrow, with $\beta > 1$. The value of k units of the good to a representative consumer is $u(k)$, where u is strictly increasing, concave, and bounded below. This infinitely lived representative consumer discounts the future with the discount factor $0 \leq \delta < 1$. We assume that the technology and preferences are such that feasible utility is bounded above.

As is standard in the theory of competitive equilibrium, the price of consumption is determined by marginal utility $p_t = u'(c_t) = u'(k_t)$. Similarly the price q_t of the durable good k_t can be computed is just the present value of future prices at which the good will sell.

$$q_t = \sum_{i=0}^{\infty} (\beta\delta)^i p_{t+i} = \sum_{i=0}^{\infty} (\beta\delta)^i u'\left(\beta^i k_t\right).$$

In other words, even when forced to compete with many other people all with access to the same reproduction technology, the initial owner of the resource will earn the present value of all future sales of the good.

Notice first that, under the usual assumption of positive marginal utility $p_t > 0$, and so it must be that $q_t > 0$ for all t. Even if the market is satiated and marginal utility falls eventually to zero, as long as $p_0 > 0$ it is still the case that $q_0 > 0$. In other words our central claim is easily satisfied: even without intellectual monopoly, and in the face of fierce competition from downstream purchasers, new ideas still command a positive price. In this special case the price of the idea decreases at a speed that depends on the ratio between $u'(\beta^t k)$ and $u'(\beta^{t+1} k)$.

Notice that technical assumptions need to be imposed to make sure that $q_0 < \infty$; to be precise, $\beta\delta < 1$ must hold, and marginal utility must decrease fast enough to guarantee that the infinite sum determining q_0 converges. Notice also that there is no upper bound on the number of units of the new good that can eventually be produced, and that there is no additional cost of making copies. Indeed, the only difference between this model and the model in which innovations are nonrivalrous is that in this model, as in reality, reproduction is time consuming, and there is an upper bound $\beta < \infty$ on how many copies can be produced per unit of time. These twin assumptions capture the observation, that nonrivalry is only an approximation to the fact that costs of reproduction are sometimes very small. This simple analysis clarifies that competitive rents can be substantial and that there is no question that innovation can occur under conditions of perfect competition.

It is worth pointing out one common error. It is natural to assume, for example, that if copyright were abolished, the price of CDs (or books, or whatever) would fall. However, a CD purchased in the absence of copyright is different than a CD purchased in the presence of copyright. The former brings with it the valuable right to make and sell copies. So when we speak of q_t as the competitive rent accruing to the sale of a CD, there is no presumption that this is less than the current sale price of a CD.

Summary

To summarize, consider the problem of innovation. After the innovation has occurred, the innovator has a single unit of the new product $k_0 = 1$ that he must sell without patent or copyright protection. In a competitive market the initial unit sells for q_0, which may be interpreted as the rent accruing to the fixed factor $k_0 = 1$ owned by the innovative entrepreneur. 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. On the other hand, introducing that first unit of the new good entails some cost $C > 0$ for the innovator. Consequently, the innovation will be produced if and only if the cost of creating the innovation C is less than or equal to the rent resulting from the innovation and captured by the fixed factor, $C \leq q_0$.

Consequences of Improving Reproduction Technology

What happens as the rate at which copies can be obtained 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 in the absence of legal monopoly protection? Conventional wisdom suggests that in this case prices will drop toward zero, and competition would necessarily fail to produce innovations. In fact, improving the copying technology leads to the most favorable circumstances for conventional wisdom. Recall that the latter is basically founded upon examination of a static model with fixed cost of invention and no cost of reproduction. What we argue next is that, even under such special circumstances, the conclusion conventional wisdom would like us to reach is not correct, at least as a matter of principle. That it fails also as a matter of facts we leave for later chapters.

Conventional wisdom fails for two reasons. First, it ignores the initial period during which, no matter how good the reproduction technology, only one copy is available. Intuitively, conventional wisdom ignores the fact that inventions usually arrive when there is some unsatisfied demand for them. If the new product is badly needed, the market value of being first, even for a relatively short period of time, may be substantial. With impatient consumers, the amount that will be paid for a portion of the initial copy (or, more realistically, for one of the few initial specimens of the new good) will never fall to zero, no matter how many copies will be available in the immediate future. This consideration has great practical relevance for markets such as those for artistic works or for medicines, where the opportunity to appreciate the work or consume the drug earlier rather than later has great value. Notice, however, that if the improvement in reproductive technology also diminishes the turnaround time for producing copies (the length of a period), then the marginal utility of consuming for a single period becomes decreasingly significant, and so this lower bound loses its economic significance. This is a point that Danny Quah has made in a continuous time context. In summary, conventional wisdom would be right in ignoring the economic value of being first if technological advances were such that the time one had to wait to purchase the good at a near zero price became negligible. Our view is that, even with the very recent and substantial advances in the information technology, this is seldom if ever the case. Claiming that it was so thirty or fifty years ago, when conventional wisdom was developed, seems rather far fetched.

Conventional wisdom also fails for a second, less apparent, reason: improving the reproduction technology may increase, rather than decrease, the rent to the fixed factor. Simply put, the creator of the idea in competitive equilibrium can claim the present value of all revenue generated by the idea. Whether price falling to zero implies revenue falling to zero depends on the elasticity of demand. If, in fact, demand is elastic, then price falling to zero implies (because so many units are sold) revenue increasing to infinity. So in this case, improved reproduction technology would increase rather than decrease the rents accruing to the innovator.

Details

These basic facts emerge clearly enough in our mathematical example. Examining the competitive rent, we see that $q_0 \geq u'(c_0)$ is bounded below by $u'(1)$ regardless of β. So, provided there is no change in the turnaround time, there is a lower bound on the competitive rent, regardless of the quality of the reproduction technology.

Turning to the issue of revenue, using the definition of q_0 we find

$$\frac{dq_0}{d\beta} = \frac{1}{\beta} \sum_{t=0}^{\infty} t(\delta\beta)^t \left[u'(\beta^t) + u''(\beta^t)\beta^t \right].$$

Take as granted the technical conditions under which this sum converges. Its value is positive or negative depending on the sign of the term within square brackets. The latter depends on the elasticity of demand or, which is the same thing, on the rate at which marginal utility of consumption decreases when consumption increases. In a dynamic context such as this one, we know that when the marginal utility of consumption decreases slowly then consumption is highly substitutable over time, and that when the marginal utility of consumption declines rapidly, future consumption is a bad substitute for current consumption. When marginal utility of consumption decreases slowly, increasing β^t decreases the competitive price at a rate lower than the growth in the quantity sold. This increases total revenue and, because total cost is the same, the rent to the fixed factor. Conversely, if marginal utility decreases rapidly, as β^t increases the price decreases faster, leading to lower rents for the innovator. Incentives to innovate are proportional to the size of the rents. Hence, when the cost of reproduction decreases, it seems that incentives to innovate will also decrease for those goods that have a rapidly declining marginal utility. Vice versa, the *incentives to innovate increase* as $\beta \to \infty$ for those goods that have a slowly declining marginal utility.

Extension of the Mathematical Example: The Interior Case

We now extend the mathematical example to consider the more general case when the good is perishable and $0 < c \leq k$ units are allocated to consumption only. The number of copies available in the following period are $\beta(k-c)$. If the good is to some extent durable, there will be ζc additional units available next period. In many cases $\zeta \leq 1$ due to depreciation, however if the good may be reproduced and consumed at the same time, we may have $\zeta > 1$. So far we have considered only the case in which $\zeta = \beta$.

To find the competitive rent, we now must consider the concave value function $v(k)$, which is the unique solution of

$$v(k) = \max_{0 \leq c \leq k} \{u(c) + \delta v(\beta k)\}.$$

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, as is the case in our original example. We now examine the case in which 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

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

It is also easy to see in the interior case that q_t must fall as the rate $1/\beta$ – this follows from the fact that one unit of capital today can be converted into β units tomorrow.

To determine the impact of changing the reproduction rate on competitive rent, we compute the derivative

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

When β is sufficiently large relative to ζ the first term will dominate the sum.

Consider the limit case in which $\zeta = 0$. In this case the rent will increase if initial-period consumption falls with β and will decrease if it rises. The mathematical nature of the result is the same as the previous case, only now the interpretation in terms of substitutability of consumption over time is transparent. If consumption is highly substitutable over time, increasing β makes the incentive for investing in future consumption stronger. The marginal utility of current and future consumption increases and, respectively, decreases only little when we shift current consumption into future consumption. The quantity available in the future, in exchange for giving up one unit today, increases with β; the investment is therefore worth taking. The rent to the fixed factor increases because such investment increases the total amount of consumption obtained from the first unit of the good, decreasing the competitive evaluation of the good, $u'(c_t)$, less than proportionally. Conversely, if there is low substitutability in consumption over time, the reduced cost of the good in future periods will not be enough of an incentive for reducing first-period consumption. Hence, as β increases, the future price of consumption decreases faster than the cost of investment increases today, leading to lower investment and higher current consumption. This decreases the total rent accruing to the owner of the initial fixed factor.

To learn more about the interior case where $\zeta < \beta$ and consumption is strictly less than capital each period, let us study the tradeoff between current and future consumption when the utility function has the CES form.

$$u(c) = -(1/\theta)(c)^{-\theta},\ \theta > -1.$$

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 $c_0 \to c < 1$. Consequently, rents from innovation fall. Still, they do not fall to zero. Competitive innovation still takes place if $p = u'(c) \geq C$.

More interesting is the case of elastic demand, where $\theta \in (-1, 0]$. This implies a high elasticity of intertemporal substitution in consumption ($\theta = -1$ corresponds to linear utility and perfect substitutability). Utility becomes unbounded above as

$$\beta \to \delta^{1/\theta}.$$

A simple calculation shows that as this limit is approached, $c_0 \to 0$ and rents to innovators become infinite.

Increasing rent with lower reproduction cost has a number of consequences. First, it means that even if the turnaround time for producing copies (the length of a period) declines as β increases, rents are nevertheless increased. When demand is elastic, increasing sales means higher revenues. Later periods, with more copies available, will therefore have more revenue, and the sooner the increase in revenue occurs, the better for the seller. When demand is inelastic, instead, increasing sales means lower revenues. Later periods, with more copies available, will therefore have less revenues, and the sooner the drop in revenues occurs, the worse for the seller.

A second observation, is that in the CES case, with elastic demand, every socially desirable innovation will eventually occur as the cost of reproduction becomes sufficiently small. Notice that, as long as the indivisible cost of innovation C is finite, one does not need demand to be elastic even at an infinite consumption level. Rather, for finite values of C one only needs demand to be elastic in some initial range $[0, \bar{c}]$, with $\bar{c}$ large enough. The fact that, when demand is elastic also a monopolist would produce ever increasing quantities of the good does not reduce the importance of this finding. Competition without intellectual monopoly leads to a socially efficient outcome whereas, as we show subsequently, even in this case intellectual monopoly brings about social inefficiency.

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. Yet the basic assumptions are simply that it takes some (small) amount of time to reproduce copies and that demand for the new product is elastic. Notice that currently accepted theories argue, as do current holders of monopoly rights, that, with the advent of a technology for cheap reproduction, innovators' profits are threatened and increased legal monopoly powers are required to keep technological innovation from faltering. In fact quite the opposite is possible: decreasing the reproduction cost can make it easier, not harder, for a competitive industry to recover production costs.

More on the Mathematics of the Cost of Reproduction

The conventional theory assumes that ideas can be redistributed at zero marginal cost. This is as absurd as the proposition that they can be distributed in unlimited quantities. In our example, we assumed that the only cost of redistribution is the fact that ideas cannot simultaneously be consumed and reproduced. How do our computations change as we vary the assumption about the reproduction cost?

Suppose that reproduction requires resources besides foregone consumption and the template good. We continue to assume that consumers have access to the same reproduction technology as the innovator. From a technical perspective, introducing a constant positive resource cost μ of reproduction is mathematically equivalent to having a lower marginal utility $u'(c) - \mu$, or to say the same thing, shifting the demand curve downwards. One key implication of doing this is that if we subtract a constant from the demand curve, for sufficiently large quantities, the willingness to pay must drop to zero, and in particular, the demand curve must eventually become inelastic.

First and significantly, the price q_0 must still necessarily be positive. No matter what the reproduction technology, there is only one unit available in the initial period, and as long as $u'(1) > 0$ this must sell for a positive price. Generally speaking, however, reproduction will absorb resources that would otherwise go to the innovator, and we generally expect a high reproduction cost to limit the competitive rents. One certain implication is that eventually enough copies will be reproduced that price drops to the resource cost of reproduction, and at this point the innovator receives no further rents. In particular, in the case of complete depreciation, the market will become satiated where $p_t = u'(c_t) = \mu$.

Because the market eventually becomes completely satiated, or put differently, because demand eventually becomes inelastic, if the capacity β increases enough, this must eventually cause q_0 to decline. However, this does not mean that improved technology for reproduction inevitably lowers the rents to the innovators. Simply put, technological change that increases β will generally lower μ, the resource cost of reproduction as well. The internet not only makes it possible to distribute more copies over a given period of time, but to do so at a lower resource cost for each unit distributed. If the resource cost falls rapidly enough, and the underlying utility function is elastic, the case is no different than discussed above, and in the special case of CES utility, the rent accruing to the innovator will in fact grow without bound as technology improves.

Note that the assumption of constant marginal cost is not crucial to this analysis. If it is increasing, this simply means that the inelastic portion of the demand curve is reached more quickly. The crucial assumption is that the innovator and consumers have access to the same technology. In many instances, the technology available to the consumers is inferior to that of the innovator. In particular, it can be costly to do reverse engineering to extract an idea from a commodity. Even in the case of literary or artistic works, they can be encrypted in such as way as to increase the cost of making copies. Costly reverse engineering will generally increase the incentive to innovate, but may lead to other distortions – for example encryption may be costly. We will examine these issues in subsequent chapters.

Implications for Competitive Rents

The conventional model proposes that ideas can be costlessly redistributed, and so competitive rents are zero, and can never be sufficient to cover the cost of innovation. Since the cost of redistributing ideas cannot possibly be zero, the conventional model can at best be viewed as a useful limiting case when the cost of redistributing ideas is small. This much we have argued and proved in the previous section.

In the case of many patentable ideas, the cost of redistribution is not small, and may well be increasing over time. Certainly the idea of how to build a wheel is much easier to communicate than the idea of how to build an atomic bomb. Basically inventions range from the trivial, such as the idea of a "single click" to buy an item on the Internet, to the complex, such as the Karmarkar algorithm for solving linear programming problems. Trivial ideas are cheap to communicate, but of course they are also cheap to create. Complex ideas are expensive to create, but they are also difficult to communicate, so they are scarce and will command a substantial premium for a long period of time. In both cases the cost of producing the ideas and the competitive rents are commensurate, and some ideas will be produced without intellectual monopoly, while perhaps others will not.

Furthermore, many of the ideas we now consider simple are, like the wheel, rather old. At the time somebody first bumped into them they were not as cheap and easy to implement, communicate and reproduce as it is often assumed in the static model of conventional wisdom. Looking at the past with the glasses of today's technology distorts one's perspective and may lead to consider all inventions of the past as cheap achievements. They are not and, as we will argue, even in the complete absence of any form of intellectual monopoly it took many decades and often centuries for the great innovations of the past to be imitated and reproduced on a large scale. This means that, if we take a historical perspective and consider useful innovations at the time their creators first introduced them, we might discover that the costs of reproduction were substantial in relation to the resources available at the time. Our previous argument implies that substantial must therefore have been also the rents accruing to the original innovator, making competitive markets a viable mechanism for delivering innovations. We leave a more detailed consideration of various specific examples the next chapter, and continue here to develop the theory of innovation without intellectual monopoly.

In the case of copyrightable creations, it can be argued that technological change – computers and the Internet – are greatly lowering the cost of reproduction, and so the limit considered in the conventional model is relevant. But, unfortunately, we have seen that the limit considered in the conventional model is not right. Competitive rents may either increase or decrease as reproduction costs decline; they may fall to zero, or increase to infinity. The striking fact is that while the impact of this technological change on competitive rents is ambiguous, the impact it has on the cost of innovation, on the size of the indivisibility (C) is not. The same technological change that is increasing reproductive capacity (β), lowering turn around time (δ^{-1}), and lowering the marginal cost of reproduction (μ), all with uncertain implications for competitive rents (q_0), is clearly and unambiguously lowering C, and it is doing so by orders of magnitude. In the case of written works, the advent of the word processor probably has not lowered C by orders of magnitude. But for music and movies, the computer has. Music editing capabilities that required millions of dollars of studio equipment ten years ago, now require an investment in computer equipment of thousands of dollars. And long before β becomes large enough to swamp markets with music and movies, authors will be able to create movies on their home computers with no greater difficulty than writing a book – and entirely without the assistance of actors, cinematographers, and all the other people that contribute to the high cost of movie making. So the implication of technological change for the cost of reproduction reinforces that of economic growth. As the cost of individual creations becomes smaller, and the audience larger, the argument in favor of intellectual monopoly fades into irrelevance.

Intellectual Property Without Intellectual Monopoly

In the setting above entrepreneurs have well-defined property rights to their innovations, individual production processes display constant returns, and there are no fixed costs and no unpriced spillover effects from innovation. Entrepreneurs also have no ability to introduce monopoly distortions into pricing. This not only provides a model for a world without intellectual monopoly, but it provides a positive theory of markets in which innovation takes place under competitive conditions. As we shall see, there are already many such markets. The question we are posing is: Why are some other markets legally removed from such beneficial competition?

Although the basic ingredients of fixed factors, rents, and sunk costs are already familiar from the standard model of competitive equilibrium, the way in which they fit together in an environment of growth and innovation is apparently not well understood. Central to the analysis is the idea that a single entrepreneur contemplating an innovation anticipates the prices at which he will be able to buy inputs and sell his output, and introduces the innovation if, at those prices, he can command a premium over alternative uses of his endowment. He owns the rights to his innovation, meaning that he expects to be able to collect the present discounted value of downstream marginal benefits. As we have shown, this provides abundant incentives for innovation.

As in theories of monopolistic competition and other theories of innovation based on increasing returns, new technologies are introduced because of the role of individual entrepreneurs in seeking out profitable opportunities. Unlike in those theories, the entrepreneur does not actually end up with a profit. Profit here is used in the technical sense academic economists attribute to it, which is not what most common people have in mind. What the layman calls profits, academic economists call rents, or competitive rents, in the context considered here. If you like to put it plainly: our competitive innovators earn competitive rents, and those rents often corresponds to lots of money. Because of competition, only the owners of factors that are in fixed supply can earn a rent in equilibrium. When a valuable innovation is introduced, it will use some factors that are in fixed supply in that period; for example: it uses the ability of the innovator to assemble or design the very first prototype of the new good, and that is most certainly in fixed supply. Those factors will earn rents. If you are good at writing operating systems code when the personal computer technology is introduced, you may end up earning huge rents, indeed. In principle, this model allows a separation between the entrepreneurs who drive technological change by introducing new activities, and the owners of fixed factors who profit from their introduction. However, it is likely in practice that they are the same people.

This is not to argue that competition is the best mechanism in all circumstances. In fact, rents to a fixed factor may fall short of the cost of producing it, even when the total social surplus is positive. Indivisibility constraints may bind. However, as we shall see below, there are many advantages to being the first mover, and competitive rents represent a lower bound on what the clever innovator can hope to get in competition with his customers, and not an upper bound.

The Mathematics of Many Ideas

We have seen that there will generally be rents for innovators, and that these rents may increase or decrease as reproduction becomes easier. But will the rents be sufficient to cover the cost of production, and which ideas will be produced when many alternative innovations are possible at any given point in time?

We now examine a world in which there are many different possible ideas, indexed by $x \geq 0$. Each idea has a social value $v(x)$ and, in the absence of intellectual monopoly, accrues a rent $q_0(x)$. We know that ordinarily the social value will exceed the rent, $v(x) > q_0(x)$. Let us order the ideas so that the more valuable ideas have lower indices; that is, $v(x), q_0(x)$ are both decreasing in x. We assume that these functions are also continuous. Ideas naturally require labor or some other resource to be produced; let us suppose that there are L units of creative labor available, and that each unit of labor has an alternative use, with a social value of $C > 0$. Each idea requires one unit of labor to be created.

There are three cases. First, it may be that $q_0(0) < C$. In this case no ideas will be produced. This may be socially optimal if $v(0) \leq C$, in which case it is not surprising that ideas with social value lower than the cost of production will not be produced. If, instead, $q(0) < C \leq v(0)$ then the indivisibility is binding even for the most valuable idea, and the competitive outcome is clearly inefficient.

If $q(0) \geq C$, let $\overline{x}$ be the idea for which the rent exactly covers the opportunity cost, $q_0(\overline{x}) = C$. If $v(\overline{x}) > C$ the socially desirable number of ideas will be greater than $\overline{x}$. If $\overline{x} < L$, then fewer ideas than is socially desirable will be produced. As we shall see later, although in both of the two preceding cases too few ideas are produced, it does not necessarily follow that allowing for intellectual monopoly is socially beneficial. There are two reasons. First, in some cases intellectual monopoly, while producing more ideas, may also produce a less socially desirable outcome, by imposing too many restrictions on the use of each idea. Second, competitive mechanisms other than the standard one considered so far may be possible, which do away with intellectual monopoly and deliver a socially superior outcome.

Finally, there is the case $\overline{x} \geq L$. In this case exactly the right number of ideas are produced, since $v(L) > C$ as well. This case is significant for several reasons. First, it is relatively robust: there a broad array of economies in which this condition will be satisfied, or nearly so. Second, it leaves no scope whatever for intellectual monopoly. In this case, the social optimum is achieved without any intellectual monopoly at all; given the many costs of monopoly, there is a strong presumption that intellectual monopoly should never be allowed in a market in which this condition is satisfied.

The point of this digression should be clear. Allowing for many potential ideas to be invented at any given point in time does not alter the fundamental intuition we developed for the case of a single innovation. The case against competition still rests on the possibility, which we discussed at length earlier on and will reconsider later, that the indivisibility involved with the sunk cost of making the initial discovery is too large relative to the marginal utility of the new good when the latter is produced at its minimum feasible quantity. Whenever the indivisibility is not binding, all ideas that have a positive social value are introduced in competitive equilibrium without intellectual monopoly.

The First-Mover Advantage

Competitive rents are the least amount that an innovator can expect to earn in the absence of intellectual monopoly. Since the innovator initially is the only one to know the idea, there are many ways to profit from this first-mover advantage. As remarkable as the phenomenon of economists who believe ideas are transmitted freely, while writing a voluminous literature on technology transfer and the cost of information, are economists who believe that innovators have no first-mover advantage, whilst writing a voluminous literature on the strategic advantages of being first. These strategic advantages are documented in most game theory textbooks: Fudenberg and Tirole is one example.

In the context of the simple model introduced here, the basic implication of the first-mover advantage is quite simple. Because, at least during the very first period, the innovator is a de-facto monopolist even without legal enforcement of intellectual monopoly, he will want to sell an initial quantity of the good which is lower than the one perfect competition dictates. This choice allows him to earn monopoly profits until imitation by his customers takes place, and provides him with more than just the competitive rents. This argument is especially relevant in the case in which demand for the good rapidly becomes inelastic which, as we have seen, is the case least favorable for competition. By exploiting the inelasticity of demand, the innovator may use the first-mover advantage to his benefit, earning substantially more than the competitive rents and, most likely, compensating for the initial sunk cost.

The most striking implications of the first-mover advantage, may well lie elsewhere. It is captured by the observation first made by Jack Hirshleifer, that the innovator, by virtue of inside information, may be able to earn vastly more than the social value of the innovation. To understand Hirshleifer's argument, consider the recent innovation of the Ginger scooter, said to revolutionize urban transportation, and grant that this unlikely prediction is actually true. How could the inventor, Dean Kamen, profit from this knowledge? There was a point in the development of the scooter at which Mr. Kamen was the only one to know that urban transportation is soon to be revolutionized, and that the automobile itself is soon to be obsolete. Rather than surrounding himself with patents, and hawking his knowledge to venture capitalists, as he did, he could simply have sold short automobile stock using whatever funds he had available to him, and leveraging to the maximum extent possible. Then, rather than developing the scooter himself, he should simply have released the blueprints to the press. As soon as the blueprints were published, the stock owning public would naturally realize that the automobile industry is on the way out, and the price of automobile stocks would plummet. Mr. Kamen, having foreseen this, and having sold short the stocks prior to publishing his blueprints, would naturally have made a killing.

In practice of course, whatever Mr. Kamen's representations to venture capitalists might have been, Ginger is unlikely to revolutionize the transportation industry, and shorting automobile stocks would have been a risky proposition. (Although in retrospect, a good decision.) But invention is a risky business in general, and the intellectual monopolist who has a valueless idea does not generally fare so well either.

There are more obvious and more common advantages of being first-mover. The primary advantage is simply that it takes time and money to reverse engineer a product. That is, in the simple models developed above, the innovator was in immediate competition with consumers. But in the short-run, reproduction and reverse engineering are expensive. Books, music, video and copyrightable items can be encrypted, and it takes time and money to crack encryption schemes. Patentable items are generally costly to reverse engineer. Moreover, the expertise that comes with being the innovator, and having been in production for longer than competitors has substantial market value. The example of Boulton and Watt after the expiration of the Watt patents is a case in point, but there are many others, such as the fact that patented drugs continue to command a substantial premium over their generic competitors, even after the patent expires. In short – even without the benefit of legal protection, the innovator certainly will enjoy a short-term monopoly.

But how is the poor inventor, working in his basement, to profit? Will not the large heartless corporations take advantage of his lack of capital to steal his idea and put it into production themselves? There is a clever scheme, explained by Anton and Yao in an article in the *American Economic Review*, showing how the inventor can avoid this. To return to the example of the Ginger scooter, Mr. Kamen could have gone to one of the automobile companies, Ford, perhaps, and shown them his blueprint for free. He would then promise to keep it secret from their competitors, but only in exchange for a substantial share in Ford Motor Co. This creates what an economist would call an incentive compatible mechanism, and what a pundit would call a win-win situation. The secret would have substantial value, since Ford would enjoy a first mover advantage. As long as Mr. Kamen asked for less than the full value of the invention to Ford, they would be happy to pay, for if he were to reveal the secret to their competitors, they would lose their monopoly profits. On the other hand, Ford would understand that Mr. Kamen, sharing in the Ford stock, would not reveal the secret to the other companies – as this would reduce the value of his stock.

Another first-mover advantage, for creative works especially, is the well-documented and strong preference for originals, signed copies and early versions that are in scarce supply, to more widely available versions. Perhaps one of the most striking examples of the phenomenon is that of the Getty Art Museum, in Los Angeles. The Getty Museum bought, at astronomical prices, a large number of very good forgeries of famous works of art. These forgeries were sufficiently good, that the experts of the Museum believed that they were originals. In other words, from the functional point of view, these works of art were the same as the originals. However, additional subtle evidence, and refined scientific testing established that indeed these works were fraudulent. Of course from the functional point of view the works were unchanged – from the viewers perspective, the painting still looked exactly the same. But the market price, once the works were clearly established as unoriginal, plummeted by orders of magnitude. Similarly, authorized copies, distinguishable from the original only by label, sell for a vastly lower price than the original. So while works of art may be currently protected by copyright – it is hard to make the case that there is any need to do so.

The preference for originals, signed or autographed copies and so forth, is just a special example of a more general phenomenon: the collateral sale. That is, a creation, while not terribly scarce in some markets, is often quite scarce in other markets, and the innovator, by virtue of being the innovator, can generally command a premium for his services in areas not directly related to his idea. Examples of this abound. In music, live performances will remain scarce, no matter what the price of electronic copies. Movies will be produced as long as first run theatrical profits are sufficient to cover production costs, and no matter how many copies are given away over the Internet for free. Books will continue to be produced as long as initial hardcover sales are sufficient to cover production costs. Substantial money is to be earned by authors or inventors by going on the talk-show circuit. Even t-shirts signed by a famous author may be enough to pay for his labor in producing his great literary work.

The greatest collateral sale of all, is, of course, the sale of advertising. Those who doubt the possibility of making a profit from giving a product away for free would do well to look into the history of the radio and television industry. How many people became fabulously wealthy from an industry that for the first 40 years of its existence had no choice but to provide its product for free? It is argued of course, that in the absence of copyright, people would simply redistribute the product with commercials removed. In the absence of technical means such as encryption, this might be possible. But of course there is nothing to prevent the creator from embedding the advertisement as an integral part of the story. Product placements are quite common in movies and television. If other advertising possibilities diminish, these will become correspondingly more valuable. There is no reason why this cannot extend to other works, such as books. While Ian Fleming did not receive payment from the Beretta Corporation for equipping his spy with a gun of that manufacture, after the books became popular, he certainly could have made a profit by auctioning off the right to the James Bond gun. In fact the Bond movies (in which he did not use a Beretta) seem to have done exactly that.

A similar possibility of collateral sale arises also in the market for patentable ideas. The inventor naturally has established special expertise in the ideas surrounding his invention. He will be in great demand as a consultant by those who wish to make use of the idea. Would not Watt have been in great demand from producers of steam engines even if he had no patent? Would Transmeta have been willing to hire Linus Torvalds at a substantial salary, had he not created Linux? Despite having given his creation away for free, and despite an apparent reluctance to profit from his fame, for example by way of public appearances, Torvalds does appear to have earned a positive return on his innovation.

Ultimately no academic work can do more than scratch the surface of the first-mover advantage: it is limited only by human ingenuity, an area in which academic economists have no special advantage. For example, profits can be made by escrowing contingent orders in advance; through serials and cliffhangers, or even by selling tickets to a lottery involving innovation as one outcome. Looking back over history we see the ingenious methods adopted by entrepreneurs in markets where indivisibilities have posed a problem. In the medieval period, the need for convoys created a substantial indivisibility for merchants that was overcome through the clever use of contingent contracts. In modern times, Asian immigrants (among other) have overcome the need for a minimum investment to start a small business by organizing small lottery clubs.

Ideas of Uncertain Value

Intellectual property absolutists, such as Jack Valenti, become extremely excited about the fact that many innovations are risky. After all, it is bad enough that competitors should be allowed to “steal” “your” creation. But if the original project is risky, they will only choose to “steal” if you are successful: few illegal copies of such great flops as *Sahara* are widely distributed on the internet.

What implication does the existence of uncertainty have for competition without intellectual monopoly? The cost of developing the innovation we have taken to be C. The amount earned in competition with many imitators we have denoted by q_0. However, if the project only succeeds with probability p, abstracting from risk aversion, the value is only pq_0. So the condition for innovation becomes $C \leq pq_0$. Naturally the lower the probability of success, the less likely the innovation is to occur. Of course, the social value of the innovation is pv, and if p is small enough $C > pv$ and it is better from a social perspective that the innovation does not occur.

In short, the uncertainty surrounding the success of an innovation changes the specific calculations of how likely it is to take place; this is true with or without intellectual monopoly. But the basic theory of competition without intellectual monopoly does not change on account of uncertainty – an uncertain outcome is equivalent to earning a lower rent, or having a higher cost. And when a lobbyist comes explaining how much more money he needs because his costs are high and his rewards uncertain…hold on to your pocket book.

Notes

Romer’s position on the role of intellectual property in growth, and the non-rivalrous nature of ideas can be found in his 1986 article in the *Journal of Political Economy* **94**, 1002-1003 “Increasing Returns and Long Run Growth” and his 1990 papers in the *American Economic Review* **80**, 97-103, “Are Nonconvexities Important for Understanding Growth?,” and the *Journal of Political Economy*, **98**, S71-S102, “Endogenous Technological

Change." Variations on this theme in the setting of monopolistic competition can be found in the work of Grossman and Helpman in the 1991 *Review of Economic Studies* **58**, 43-61, "Quality Ladders in the Theory of Growth." These ideas build on the earlier ideas of Alwyn Young, and especially the post-war work of Kenneth Arrow in "Economic Welfare and the Allocation of Resources for Invention," in *Rate and Direction of Inventive Activity: Economic and Social Factors*, 609, 617 (National Bureau of Economic Research ed., 1962). This ideas are developed in further detail by Karl Shell in 1996 in the *American Economic Review* **56**, 62-68, "Toward a Theory of Inventive Activity and Capital Accumulation" and in 1967 in "A Model of Inventive Activity and Capital Accumulation" in *Essays on the Theory of Optimal Economic Growth* (K. Shell, ed.), Cambridge, Massachusetts: MIT Press, 67-85.

A more extreme view is that of Schumpeter, who in his 1911 *The Theory of Economic Development*, translated into English in 1934 (New York: McGraw Hill) who celebrates monopoly. This theme is elaborated by in the modern economics literature by Aghion and Howitt in 1992 in *Econometrica* **60**, 323-351, "A Model of Growth through Creative Destruction."

The extensive microeconomics literature on patents generally departs from the assumption that innovation will not take place without a patent, and inquires into the optimal length and breadth of patent protection. Good examples can be found in the work of Gilbert and Shapiro, or Gallini and Scotchmer.

The modern view is exposited by the two of us in 2002 in the *American Economic Review* "The Case Against Intellectual Property" and by Hellwig and Irmen in 2001 in the *Journal of Economic Theory* **101**, 1-39, "Endogenous Technical Change in a Competitive Economy" available online at http://www.vwl.uni-mannheim.de/hellwig/irmen/radner.pdf.

In dealing with the mathematical details, note that in a model of growth, the assumption of CES utility need only hold above a certain minimum subsistence level of consumption. The Quah 24/7 model appears in his working paper "24/7 competitive innovation" available online at http://econ.lse.ac.uk/staff/dquah/currmnu1.html.

Much discussion of the first-mover advantage in game theory can be found in Fudenberg and Tirole's book *Game Theory*.

The Hirshleifer model appears in "The Private and Social Value of Information and the Reward to Inventive Activity" in *The American Economic Review*, Vol. 61, No. 4. (Sep 1971), pp. 561-574. The profit sharing scheme for protecting innovators can be found in "Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights" by James J. Anton and Dennis A. Yao in AER 84 p. 190-209, 1994.

Information about Dean Kamen and Ginger was widely reported in the press during 2001. Information about the Getty Museum, Ian Fleming, Linus Torvals and Jack Valenti are surely available somewhere.

This chapter benefited from comments and careful reading by John Gallup.