The Economic Structure of Intellectual Property Law

William M. Landes

Richard A. Posner

The Belknap Press of
Harvard University Press
Cambridge, Massachusetts, and London, England | 2003

Copyright Act of 1976 and the further lengthening of the term by the Sonny Bono Copyright Term Extension Act (see Chapter 8). Because of discounting to present value, extensions of the copyright term beyond twenty or twenty-five years have little incentive effect (and thus do not bring forth a significant number of additional expressive works to enrich the public domain when the copyright on those works expires), but greatly diminish the size of the public domain, especially since all extensions of the copyright term have been applied to existing as well as new works. We address the issue of optimal duration in Chapter 8; Chapter 4 focuses on issues of scope.

A final caveat: throughout this chapter our emphasis has been on the effect of copyright on the production of expressive works. In the terms introduced in Chapter 1, the emphasis is on the dynamic as distinct from the static benefits and costs of copyright. But we shall see in Chapter 8 that there may be static benefits of copyright as well. Copyright may correct certain congestion externalities in the market for expressive works. These benefits, which previous scholarship on copyright has largely overlooked or denied, must be kept in mind in any overall assessment of the social value of copyright.

3

A Formal Model of Copyright

As explained in the last chapter, while standard economic models of copyright emphasize the incentive-access tradeoff, we emphasize in addition the tradeoff between the incentive and cost-of-expression effects of varying the level of copyright protection. In order to incorporate this insight in a tractable formal model, we make several simplifying assumptions: that creators and copiers produce quality-adjusted copies that are perfect substitutes,[1] that demand is not subject to uncertainty, that the cost of expression is the only fixed cost of an expressive work, and that the marginal costs of creators, though not of copiers, are constant. We shall let p denote the price of a copy, $q(p)$ the market demand for copies of a given work, x and y the number of copies the creator and the copiers produce, respectively (so $q = x + y$), c the creator's marginal cost of a copy, and e the cost of expression. We denote the level of copyright protection by $z \geq 0$, so that $z = 0$ signifies no copyright protection and $z = 1$ signifies complete protection—no copying is permissible without the copyright holder's consent. The amount of copyright protection depends on such things as how alike two works must be before infringement will be found, the elements in a work that are protected, the duration of protection, and the efficacy and cost of enforcement. We subsume all these factors in our single index of copyright protection, z.

We assume that copiers, like fringe firms in a market with a dominant firm, supply copies up to the point where price equals marginal cost and that their marginal cost increases (not necessarily steeply) as both the number of copies and the level of copyright protection increase.[2] Given our earlier assumption that the author's[3] marginal cost (c) is constant, increasing marginal cost for

1. A copier might, of course, produce a copy only half as good as one produced by the author. In that event, on a quality-adjusted basis two copies made by the copier would be weighted as one.

2. We explained the significance of this assumption in Chapter 2.

3. Remember that we are using "author" and "creator" interchangeably and ignoring the difference between author and publisher.

copiers is a necessary assumption; otherwise copiers would produce all copies, in which event the work would not be created, or no copies, in which event the degree of copyright protection would not present an interesting question.

More important, it is *realistic* to assume that copiers will have increasing marginal costs. Recall that the copying that takes place at a given level of z is lawful. Some of it will be by consumers (for example, home taping of television programs) and some by producers who incorporate the author's work into their product (for example, fair use copying). The higher z is, the less the amount of such lawful copying. At a given level of $z < 1$, however, there will be some types of copying that require consumers and producers to use only a small amount of their own resources. They will be able to free ride on the author's work, and so the cost of copying will tend to be low. Other types of copying will be more costly, and here free riding will be less important. Such differences should generate differences in the cost of copying among copiers and so lead to rising marginal costs for the copiers as a group (rising because if demand falls, more of it will be supplied by the copiers having the lowest marginal costs).

The copiers' supply curve can thus be written as

$$y = y(p, z) \tag{1}$$

with $y_p > 0$ and $y_z < 0$.[4] The author's profits are

$$\Pi = (p - c)x - e(z), \tag{2}$$

and substituting for x we have

$$\Pi = (p - c)[q(p) - y(p, z)] - e(z), \tag{3}$$

4. That is, an increase in market price, or decrease in copyright protection, will evoke additional supply. We can derive the copiers' supply curve from their cost as follows. Denote the total cost of copying for copiers $M = M(y, z)$ where $M_y > 0$ (= marginal cost), $M_{yy} > 0$, $M_z > 0$, $M_{zz} \geq 0$, and $M_{yz} > 0$. That is, marginal cost is positive and increases with the number of copies. We assume that z increases both the total and marginal cost of copies (M_y) because, as z increases, the amount of protected material in a given work will rise, so that copiers must add more of their own material or make greater alterations in the copy to avoid infringement. This factor will tend to make copying more costly. To simplify, we assume that $\partial M_{yy}/\partial z = 0$—that is, that the rate of change in marginal cost is independent of z. Since copiers operate where $p = M_y(y, z)$, we have $y_p = \partial y/\partial p = 1/M_{yy} > 0$ and $y_z = \partial y/\partial z = -M_{yz}/M_{yy} < 0$. Note that y_z denotes the shift to the left in the copier's supply curve as z increases, so that at each price copiers make fewer copies.

where $e(z)$ denotes the author's cost of expression, which is higher the greater copyright protection is.[5]

Let the author's gross profits, R in our notation, equal his revenue from selling copies minus the cost of making those copies, or $(p - c)x$. We show later that R increases as z increases. The author will create a work only if

$$R \geq e(z), \tag{4}$$

since otherwise his profits (equation (2)) would be negative.

Let N equal the total number of (equivalent) works that are created. Our assumption that the cost of expression, $e(z)$, will differ among authors—some authors will be more efficient at creating equivalent works and so their costs will be lower than those of other authors—implies that with free entry of authors into the business of creating new works, N will rise until the cost of expression of the marginal author equals R. The supply of works will equal

$$N = N(R, z), \tag{5}$$

where $N_R > 0$ and $N_z < 0$.

The net effect on N of an increase in copyright protection (z) depends on the balance between two effects because the increase leads to both a movement up the supply curve of works (as R increases) and an upward shift of the supply curve as z drives up the cost of expression. Thus $dN/dz = N_R(dR/dz) + N_z$. At low levels of z, the revenue-enhancing effect of limiting copying by free riders should dominate, so that $dN/dz > 0$. When z is very low, few or no works may be created, since free riding by copiers may prevent any author from covering his cost of expression.[6] So N will increase as z increases, at least up to some level, say $\bar{z}$. Beyond $\bar{z}$ we assume that increases in the cost of expression to marginal authors will dominate, so that the number of works will begin to fall. That is, $dN/dz > 0$ for all $z < \bar{z}$, $dN/dz = 0$ at $\bar{z}$, and $dN/dz < 0$ for all $z > \bar{z}$.

5. Our model is similar to one used by Salop and Scheffman to analyze how a dominant firm selects strategies that raise both its and its rivals' costs. See Steven C. Salop and David T. Scheffman, "Cost Raising Strategies," 36 *Journal of Industrial Economics* 19 (1987). In our model copyright protection is like a strategy that raises both the rivals' (copiers') marginal cost and the dominant firm's fixed cost (the creator's cost of expression).

6. If, however, the copier's marginal cost is much higher than the original author's marginal cost, copyright protection may not be necessary for the author to be able to cover his full cost of expression, as we know from our earlier discussion. In such a case the principal effect of increasing z, even at low levels, will be to raise the cost of expression and thus lower N. In our formal model we assume away conditions that would make the optimal amount of copyright protection equal zero.

The intuition behind these results is straightforward. Some copyright protection is necessary to generate incentives to incur the costs of creating easily copied works. But too much protection can raise the costs of creation to a point at which current authors cannot cover their costs even though they have complete copyright protection for their own originality. The key issue is how the level of protection, z, here modeled as a single index, is set along several dimensions that include withholding protection from ideas as opposed to expression, giving copyright holders rights over derivative uses, and permitting unauthorized copying that satisfies fair use criteria.

Optimal copyright protection depends not only on the number and cost of original works but also on the number and cost of copies of each original work. Once created, an expressive work can be exploited in many ways, and they need not be mutually exclusive. The author (or publisher, remember) of a novel may sell copies, sell to a magazine prepublication rights to publish selections, and license derivative works such as a play, musical, movie, translation, or condensation. Similarly, he may license the characters in his novel for a comic book or television series or for a line of clothes. In the next chapter we distinguish between copiers who make identical copies of a work and those who create derivative works, since the former have a greater impact on the incentive to create the original work. For now, we treat all different ways to exploit a work identically and call them the making and selling of copies.

Figures 3.1 and 3.2 may help to convey an intuitive understanding of how copyright protection (z) simultaneously determines the price of a copy (p), the number of copies produced by the creator of the expressive work (x), the number of unauthorized copies (y), the economic returns to creating a work (R), and the total number of works created (N).

Figure 3.1 illustrates the market for copies. The demand curve for copies produced by the creator of the expressive work is derived by subtracting the copiers' supply curve ($y = y(p, z^0)$) from the market demand for copies. The creator of the expressive work then sets marginal cost (c) equal to the marginal revenue from the derived demand curve. This yields a price of p^0, quantities of copies produced by the creator and copier of x^0 and y^0 respectively, and total number of copies of q^0 ($= x^0 + y^0$). In equilibrium, the creator of the expressive work earns gross profits shown by the shaded area labeled R^0. Notice that the copiers' marginal cost or supply curve depends on the level of copyright protection—as z increases, $y(z)$ rotates to the left—which in turn affects prices, output, and the gross profits earned by the copyright holder: an increase in z increases p, x, and R and lowers y and q.

In the market for the creation of new expressive works, depicted in Figure 3.2, the number of works (N) increases with an increase in copyright protection to the point at which the cost of the marginal author's expression, which

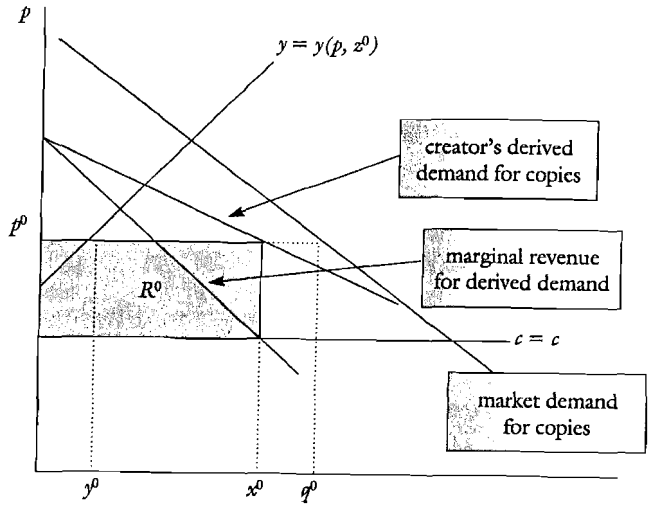

Figure 3.1. The market for copies

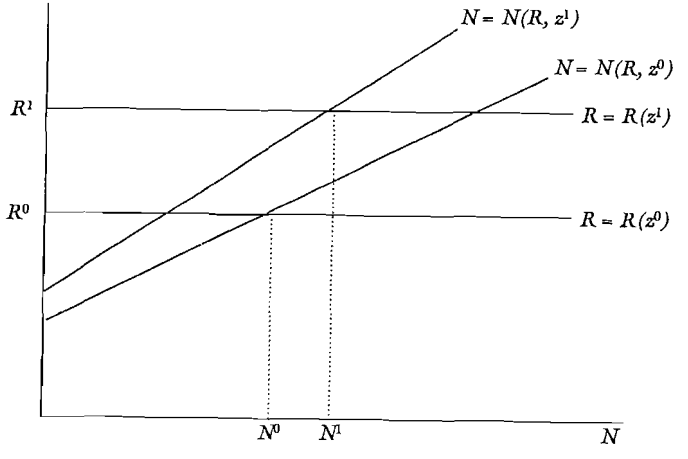

Figure 3.2. The market for new works

we assume differs among authors and is therefore increasing in N, equals his return. Assume that copyright protection is initially set at z^0, which in turn determines the equilibrium values in Figure 3.1 and in particular the gross profits (R^0) of creators of expressive works. From Figure 3.2 we can then determine the equilibrium number of new works; it is N^0. If copyright protection increases, the R curve will shift upward to R^1 but so will the supply curve N of new works because expanding copyright protection by diminishing the public domain increases the cost of creating new intellectual property. Although N is shown as increasing from N^0 to N^1 as copyright protection expands, the net effect of an increase in z on N actually is ambiguous—and for the additional reason that an increase in N brought about by an increase in z has a feedback effect by increasing competition in the market for expressive works and thus dampening the effect of the increase in z on the copyright holder's revenue.

Below we solve for the level of z that maximizes social welfare, but the considerations determining this maximum can be seen in Figures 3.1 and 3.2. In the market for copies (Figure 3.1), an increase in z produces a greater deadweight loss because the price of a copy increases and the number of copies sold decreases. The increase in z also increases the creator's gross profits, R, which in turn leads to both a movement up and a shift to the left in the supply curve of new works in Figure 3.2. The first effect creates social welfare in the form of producer surplus, while the second reduces producer surplus. Overall social welfare is maximized when the marginal benefit of increasing z in higher producer surplus exactly balances the reduction in welfare in the market for copies plus the reduction in producer surplus in Figure 3.2 as the supply curve of new works shifts upward.

The Price of a Copy

The author will choose the price that maximizes his profits in equation (2). This requires that p satisfy

$$[q(p) - y(p, z)] + (p - c)(q_p - y_p) = 0, \tag{6}$$

which can be rewritten as

$$p\{1 - F/[\varepsilon^d + \varepsilon^s(1 - F)]\} = c, \tag{7}$$

where F is the fraction of copies made by the author, $1 - F$ the fraction made by copiers, ε^d the elasticity of demand for copies, and ε^s the elasticity of supply of the copiers ($\varepsilon^s = y_p(p/y)$). The price per copy will be greater the less elastic the demand for copies, the less elastic the copiers' supply curve, and the larger the author's share of copies relative to that of copiers, which in turn will be larger the lower the author's cost of making copies relative to that of copiers.[7]

We can determine the effect on price of changes in the level of copyright protection (z), and in the author's marginal cost of copying, by totally differentiating p in equation (6) with respect to z and c. This yields

$$dp/dz = y_z/S > 0, \tag{8}$$

$$dp/dc = (q_p - y_p)/S > 0, \tag{9}$$

where S equals $\partial^2\Pi/\partial p^2$, which is negative from the second-order condition for profit maximization. Increases in z and c increase the price of a copy and reduce the total number of copies sold—provided, of course, that copiers' output is still positive. If not, increases in z will have no effect on the price and output or the number of copies; the author will be a monopolist and will not need any copyright protection.

We are also interested in the effect of changes in z on the author's gross profits (that is, before deducting the cost of expression) and on the number of copies made by author and copiers (again assuming that copiers have a positive output). The change in gross profits (R) from a small change in z is given by

$$dR/dz = -(p - c)y_z > 0.^8 \tag{10}$$

Since y_z in equation (8) is the reduction in the quantity supplied by copiers as z increases (holding p constant), the change in the author's gross profits for a small increase in z will equal the difference between price and the author's marginal cost multiplied by the increased number of copies he supplies, an increase that in equilibrium will just match the reduction in copies supplied by copiers.

Although the author's gross profits will increase with greater copyright protection until copiers cease making copies—after which additional copy-

7. We assume that the second-order condition for a maximum is satisfied—that is, that

$$\partial^2\Pi/\partial p^2 = 2(q_p - y_p) + (p - c)(q_{pp} - y_{pp}) < 0.$$

8. dR/dz equals

$$d(p - c)x/dz = dp/dz + (p - c)\{q_p(dp/dz) - [y_p(dp/dz) + y_z]\}.$$

Collecting terms yields

$$d(p - c)x/dz = dp/dz[x + (p - c)(q_p - y_p)] - (p - c)y_z.$$

Since $x + (p - c)(q_p - y_p) = 0$ from the first-order profit-maximizing condition, the first term in the above expression vanishes, leaving the expression in the text.

right protection can yield no benefit since there are no more competitors to exclude—net profits need not rise. The cost of expression to authors of copyrighted works increases as copyright protection increases, because of transaction costs, acquisition costs (the license fee charged by the owner of a copyrighted work that the new producer wants to incorporate in his work), and substitution costs (finding some equivalent in the public domain to the copyrighted input that the new producer would most like to use). And so the less material an author (not a copier) can borrow from other copyright holders without infringing their copyrights unless he has licenses from them, the greater will be his cost of expression. The change in net profits from increases in z will be positive or negative depending on whether

$$-(p - c)y_z - e_z < \text{or} > 0. \tag{11}$$

The sign of inequality (11) bears on the earlier question of whether an increase in copyright protection will increase or decrease the number of new works created. A positive sign for the marginal work (or author) means that an increase in z increases gross profits by more than the cost of expression, so net profits will rise and the number of works will increase. A negative sign means that greater copyright protection will reduce the number of works.

We speculated earlier that at low levels of z the revenue-enhancing effect would dominate while at higher levels the cost-enhancing effect would dominate. Inequality (12) enables us to be more explicit about the factors that affect the relation between z and the number of works. Since gross profits equal the cost of expression for the marginal author, inequality (11)—the condition for whether z increases or decreases the number of works—can be rewritten in percentage terms:

$$-\tilde{y}_z(y/x) - \tilde{e}_z < \text{or} > 0, \tag{12}$$

where $\sim$ denotes percentage change brought about by a change in z. This expression is more likely to be negative the smaller the copiers' share relative to the author's share (that is, the smaller is y/x). Since the copiers' share will fall and the author's rise as z increases, inequality (12) is more likely to be negative at higher than at lower levels of copyright protection. So, consistent with our earlier conjecture, the revenue-enhancing effect of increasing copyright protection diminishes as the level of protection increases.

The percentage change in the copiers' supply brought about by a change in the level of copyright protection, $\tilde{y}_z$, will be greater the larger the increase in the copiers' marginal cost as z increases and the lower the rate of increase in marginal cost with respect to a change in the number of copies—that is, the lower their supply elasticity.[9] This has two implications. The first is that the more difficult it is for copiers to avoid infringing the author's copyright by substituting other inputs for the protected part of the author's work because the protected part is bigger, the greater will be the increase in the copiers' marginal cost. The greater, therefore, will be the increase in the author's gross profit, and so the more likely is the number of works to increase as copyright protection expands. If copiers produce only exact copies or, equivalently, slavish imitations, there will be, by definition, no other inputs to substitute for the author's work, and therefore an increase in z will tend to have a large positive effect on the copiers' marginal cost curve and so on the number of works created.

Second, the smaller the difference in efficiency or cost of copying among copiers (which depends in turn on the similarity of the uses that copiers make of the author's work), the more elastic the copiers' supply or marginal cost will be and the larger therefore will be the increase in the author's gross profit as z increases. This, too, makes it more likely that expanding copyright protection will increase the number of works created. Alternatively, if copiers use the author's work in diverse ways, the marginal cost of copying is likely to be less elastic and so an increase in copyright protection will have a smaller effect on the author's gross profits.

What happens to the number of copies produced by copiers and by the author as the level of copyright protection rises? Since price will rise, the total number of copies will fall. The change in the copiers' output (y), however, will depend on the net effect of two offsetting effects. As z rises, the copiers' supply curve will shift to the left ($y_z < 0$), reducing y. But the increase in p will lead to a movement up the supply curve, increasing y.

Regarding the number of copies sold by the author (x), and recalling that $x = q - y$ and that an increase in copyright protection raises price and lowers the total number of copies sold, the author will sell more copies only if y declines by more than the reduction in q. That is indeed the most likely outcome, however. Since an increase in z raises the residual demand faced by the author, he would normally be expected to sell more. But if the elasticity of the residual demand curve declines sufficiently as it shifts outward, the author may produce less at the new equilibrium price. This is simply an illustration of the well-known proposition that an increase in demand may reduce the optimal output of a monopolist if the elasticity of demand declines sufficiently with the increase.

9. Recall that $-y_z = M_{yz}/M_{yy}$ where M_y denotes the copiers' marginal cost. Hence $-y_z$ will tend to be greater the greater the increase in the copiers' marginal cost as z increases and the smaller the increase in their marginal cost as y increases (that is, the flatter or more elastic the copiers' marginal cost curve).

Welfare Effects of Copyright Protection

To model the effects of copyright protection on economic welfare, let w equal the standard measure of economic welfare (the sum of consumer and producer surplus) in the market for copies of a single work before the cost of creating the work is deducted:

$$w = \int_{p^*}^{\infty} q(p)dp + (p^* - c)[q(p^*) - y(p^*, z)] + \int_{p0}^{p^*} y(p, z)dp \quad (13)$$

The first term is consumer surplus at p^* (the profit-maximizing price set by the author), the middle term is the author's gross profits, and the last term is the copiers' profits.[10]

Net welfare equals $w - e(z)$, where $e(z)$ is the cost of creating the particular work and is a function of the scope of copyright protection. The change in net welfare with respect to a change in z equals

$$\partial[w - e(z)]/\partial_z = (p^* - c)[q_p(dp^*/dz) - [y_p(dp^*/dz) + y_z]] + \int_{p_0}^{p^*} y_z dp - e_z \neq 0. \quad (14)[11]$$

This complicated expression has a simple interpretation. The first term is the change in the author's surplus from a change in the scope of copyright protection. It depends on the difference between price and the author's marginal cost and on the change in the number of copies he sells; for the term in brackets is merely the difference between the change in total copies and the number of copies sold by copiers. Normally the author will sell more copies when z increases because the copiers' marginal costs will rise.

Notice that at the margin copiers generate no consumer or producer surplus, because they equate marginal cost to price. As for the last two terms in equation (14), $\int_{p_0}^{p^*} y_z dp$ is negative because an increase in z increases the

10. Notice that p_0 is the minimum price at which copiers are willing to produce a copy. Since we assume that copiers incur no fixed costs, the number of copies at p_0 is zero—that is, $y(p_0, z) = 0$.

11. From equation (13) we have

$$\partial w/\partial z = -q(p^*)(dp^*/dz) \quad \text{(i)}$$
$$+[q(p^*) - y(p^*, z)](dp^*/dz) + (p^* - c)[q_p(dp^*/dz) - (y_p(dp^*/dz) - y_z] \quad \text{(ii)}$$
$$+ y(p^*, z)(dp^*/dz) - y(p^0, z)(dp^0/dz) + y_z dp \quad \text{(iii)}$$

where (i) denotes the reduction in consumer surplus ($dp^*/dz > 0$); (ii) the change in the author's profits, which depend on both the change in price and the change in the number of copies he produces (which may be positive or negative); and (iii) the change in the copier's surplus, which depends on the increase in price, the increase in his costs, and the change in the number of copies he produces. Combining the terms in $\partial w/\partial z$ yields equation (14).

total cost to copiers of the copies they produce; $-e_z$ is also negative because the cost of expression increases with the amount of copyright protection.

An increase in copyright protection is likely to reduce the welfare benefits (consumer plus producer surplus) generated by a given work, assuming it will be created. Both the increase in the cost of creating the work and the increase in the cost to copiers reduce welfare, and only rarely will these increases be offset by cost savings resulting from the shift in producing copies from copiers to the author, a shift that will be larger the lower the author's marginal cost relative to that of the copiers. For the cost savings are obtained only on the additional units produced by the author, while the cost increase affects all copies produced by copiers plus the cost of expression.

Total welfare, however, depends on the number of works created as well as on the consumer and producer surplus generated by a given work assuming it is created; and the number of works may rise as copyright protection expands even though welfare per work falls. The traditional analysis emphasized the tradeoff between the benefits of copyright protection in encouraging the production of works and the losses from reducing access to the works by consumers. If one defines "access" as the sum of consumer and producer surplus generated by a single work, access is indeed likely to fall as copyright protection increases. But it falls because of factors—the increase in copiers' costs ***and*** in the cost of expression—that are ignored in the traditional view. That view stresses losses to consumers from higher prices—a factor that drops out of our analysis.

Let total welfare equal

$$W = W[N, w, E(N, z)]. \quad (15)$$

W will be an increasing function of both N, the number of (equivalent) works created, and w, the consumer and producer surplus per work before deducting the cost of creating the work, and will be a decreasing function of E, the total costs of creating works (including the cost of administering and enforcing the copyright system). In turn, E will be an increasing function of both N and z (that is, $E_N > 0$ and $E_Z > 0$).[12] We assume for convenience that equation (15) can be rewritten as

$$W = f(N)w - E(N, z), \quad (16)$$

12. Note that $E_{NN} > 0$ because authors will differ in the costs of creating works, and as the economic return from creating works increases, higher-cost authors will find it economical to create works. $E_{Nz} > 0$ because increasing copyright protection will raise the cost to all authors of creating works. Administrative and enforcement costs are likely to rise both with the number of works created, holding constant z, and with the level of copyright protection, holding constant N, since more works will be registered and more infringement suits brought. One possible offset, however, is that an increase in copyright protection will deter some infringers. In that event, the number of suits may fall despite the greater incentive to pursue infringement claims as z increases.

where $f_N > 0$ and $f_{NN} < 0$—that is, there is diminishing marginal utility as the number of works created increases.

Maximizing W with respect to z yields

$$\partial W/\partial z = f_N N_z w + f(N) w_z - (E_N N_z + E_z) = 0, \tag{17}$$

or, equivalently,

$$N_z(f_N w - E_N) = -f(N) w_z + E_z, \tag{18}$$

where $N_z = (\partial N/\partial R) R_z + (\partial N/\partial z)$ and $w_z = (p^* - c)(dx/dp)(dp/dz) + \int_{p^0}^{p^*} y_{zdp}$ (see equation (14)).[13] We denote by z^* the level of z that maximizes W. The right-hand side of equation (18) will be positive at z^* in the typical case because an increase in z will lower producer and consumer surplus per work (that is, w_z is negative) and raise the cost of expression for all works and increase administrative and enforcement costs ($E_z > 0$).[14]

N_z measures the response of the number of works created to an increase in copyright protection. As we saw earlier, it can be either positive or negative. However, when z is set optimally, N_z will be positive. For suppose that N_z were negative at z^*. Since the same level of N could be attained at a lower z (because N increases initially and then falls as z rises), a lower z would yield a higher level of W. Not only would $E(N, z)$ be lower (since it is a positive function just of z when N is unchanged, and z would now be lower), but w (consumer and producer surplus per work before deducting the cost of expression) would be higher at a lower z for reasons explained in the previous section.

We can therefore eliminate from our analysis levels of z at which $N_z < 0$. The only exception would be where w fell as z fell—that is, where the loss in producer surplus from substituting copies made by copiers for those made by authors exceeded the reduction in copiers' costs as z fell, provided that, in addition, this effect was large enough to offset the reduction in E as z fell. We showed earlier, however, that w is likely to rise as z falls. Moreover, if total welfare were maximized when N_z was negative, this would turn the traditional rationale for copyright protection upside down. Instead of encouraging the production of works, copyright would discourage them in equilibrium, and instead of reducing access it would increase access (defined, as before, in terms of welfare per work).

Another consideration, not captured in our formal model but working in the same direction, is that as N rises a point may be reached at which any further increase will raise each author's cost of expression and hence E_N, the marginal cost of expression. With more and more copyrighted works, the amount of public domain material—unappropriated materials suitable for inclusion in a new work—will fall. It will then cost more to create a new work. This problem would be particularly serious if "ideas" (in the sense used in copyright law) as well as expression were copyrightable. They are not, as we shall see in the next chapter, and our explanation will be based on the relation between z and N developed in our formal model.

Several implications of our formal model should be noted, some of which will figure in subsequent chapters:

1. At z^*, the amount of producer and consumer surplus per work (w) weighted by f_N must exceed the cost of creating the marginal work;[15] otherwise the left-hand side of equation (18) would be negative. This implies that the optimal amount of copyright protection is greater for classes of work that are more valuable socially (that is, the higher w is relative to the cost of creating the work). The left-hand side of equation (18) would rise initially, relative to the right-hand side, requiring an increase in z to restore equilibrium. Contrary to our model, the law has resisted efforts to vary the term or otherwise alter the amount of copyright protection depending on the perceived social value of different classes of expressive work. This may be the correct second-best solution, however, because of the invitation to politicization of the copyright process and resulting rent seeking that would be tendered if copyright protection varied among classes of work. The more heterogeneous the class of persons affected by a law, the more difficult it is for them to organize a politically effective interest group to make the law more favorable to them.

2. Optimal copyright protection requires that z^* be set *below* the level that maximizes the number of works created. The latter would require that $N_z = 0$ (assuming that N increases initially and later decreases as z increases), which would make the left-hand side of equation (18) zero. To put this differently, strengthening copyright protection beyond z^* would increase the incentive to create more works ($N_z > 0$) but would not be worth the costs in reduced welfare per work, the higher costs of expression (for works that would have been created anyway at a lower value for z), and the greater administrative and enforcement costs. This conclusion formalizes the tradeoff between the cost of expression, which copyright protection increases, and the incentive effect of copyright in encouraging the creation of expressive works.

3. It follows from equation (18) that the greater the responsiveness of N to

13. We assume that the second-order condition for a maximum is satisfied, that is, that $\partial^2 W/\partial z^2 < 0$.

14. The right-hand side of equation (18) could be negative in the unlikely event that w increased and enforcement costs decreased as z increased, and these changes more than offset the increased cost of expression. In that event, z^* would be at a level sufficient to eliminate all

15. E_N also includes the incremental cost of administering and enforcing the copyright system that results from an increase in N. Note that since $E_{NN} > 0$, w will also exceed the full cost of creating each work at z^*.

an increase in z (that is, the greater N_z at each level of z), the greater the optimal value of copyright protection must be to reach equilibrium. In turn, N_z will be greater (as z increases) the greater the increase in gross profits (R) (which is greater the greater the difference between p and c and the bigger the reduction in copies made by copiers as z increases), the smaller the increase in the cost of expression for the marginal author, and the smaller the rate of increase in the marginal cost of expression as N_z increases and hence the smaller the difference among authors in the cost of creating works. In other words, optimal copyright protection will tend to be greater the more responsive the supply of new works is to increases in such protection, and this responsiveness will depend on both the costs of creating new works and the costs of producing unauthorized copies. The higher the costs of new works and the lower the costs of producing unauthorized copies, the greater the optimal scope of copyright protection.

4. We know that the optimal extent of copyright protection tends to rise with the value of a work (w) and that w will be greater the greater the demand for the work and the lower the marginal cost of making copies. Hence if rising incomes and technological advances enlarge the market for a representative work and the cost of copying declines, copyright protection should expand. This suggests a possible efficiency explanation for why copyright protection has, indeed, expanded over time. But it is highly conjectural because the model does not specify the optimal extent of that expansion, which may have gone too far, as subsequent chapters will suggest.

5. Suppose that w falls only slightly as copyright protection expands. Then the right-hand side of equation (18) will be smaller and the optimal level of copyright protection will rise. Put differently, the less that welfare per work is reduced by copyright protection, the higher will be the optimal level of that protection because an expansion in copyright protection increases the number of expressive works.

6. The more the cost of expression rises as z increases (that is, the greater is E_z), the smaller will be the optimal amount of copyright protection. This suggests that if it is feasible to differentiate in infringement proceedings between individuals who make literal copies and those who use copyrighted material to create derivative works by adding new expression to the copyrighted original, there should be broader copyright protection against the former group than against the latter—and there is.

7. Finally, and obviously (but sensibly), the formal analysis implies that the lower the cost of administering and enforcing a copyright system and the more responsive authors are to pecuniary incentives, the greater will optimal copyright protection be.

4

Basic Copyright Doctrines

The model developed in the preceding two chapters can help explain a number of features of copyright law. We examine several of them here and others (primarily fair use, discussed briefly in this chapter, and the length of the copyright term) in subsequent chapters.

Copying versus Re-Creation

We begin with the nature of the protection that a copyright gives its owner. In contrast to a patent or trademark, a copyright protects only against copying; unintended re-creation of the copyrighted work ("independent," that is, inadvertent, duplication) is not actionable.[1] Liability for intentional copying is, however, strict in the sense that it is no defense that the copier reasonably believed that the work was in the public domain.

Economic analysis suggests two reasons why inadvertent duplication is not actionable. The first is the added cost to an author of checking countless copyrighted works to avoid inadvertent duplication. In terms of our formal model, this cost (if actually incurred—a qualification the significance of which will appear shortly) would increase $e(z)$ (the cost of expression as a function of copyright protection) and lower social welfare, because both net welfare per work ($w - e(z)$) and the number of works created would fall. True, the author's gross revenues might rise if the reduction in the amount of

1. Although we can think of an exception: an art critic meticulously analyzes and inspects a famous painting of, say, Niagara Falls. By microscopic study of the brushwork, pigments, and other design elements, he is able to determine the exact place and time of day and of the year at which the painting was made and to compile a set of instructions that if followed exactly would enable an amateur to make an identical painting of Niagara Falls without ever having seen the famous painting. Even though the amateur had not knowingly copied the painting, this would not be independent duplication; he would be an infringer and the art critic a contributory infringer.

8

The Optimal Duration of Copyrights and Trademarks

In this chapter we undertake a critical examination of the widely accepted proposition that economic efficiency requires that the copyright term be limited[1] and glance at the contrasting case of trademark protection, which as explained in the preceding chapter has no time limit as such although it can be lost by abandonment or laches or by becoming a generic name. The reader may be surprised to find us toying with the idea of "perpetual" copyright, in light of our emphasis on uncertainty that the existing scope of intellectual property rights can be justified economically. But the tradeoff we focus on in this chapter is not life plus seventy years versus forever but life plus seventy years versus short fixed terms renewable as many times as the copyright owner wants if he is willing to pay a renewal fee (which may be substantial) every time. The result might be a larger public domain, and in particular fewer restrictions on copying most works created recently, than under the current system.

We do not consider the possible application of the approach to software copyrights. Indefinite renewal might enable software producers to impede competitors' software development, an unlikely prospect with regard to other types of copyrighted work. In any event, software is excluded from our empirical analysis, which (as will become clear) is necessarily limited to experience under the pre-1976 copyright regime, when there were no software copyrights.

Introduction

The first federal copyright statute, enacted in 1790, specified an initial term of fourteen years plus a renewal term of the same length, provided the author was still living at the end of the initial term. The initial term was lengthened to twenty-eight years in 1831 and the renewal term to twenty-eight years in 1909, forty-seven years in 1962, and sixty-seven years in 1998. The Copyright Act of 1976 switched from a fixed to a variable but still limited term

1. See, for example, Brief of George A. Akerlof, et al., as Amici Curiae in Support of Petitioners in Eldred v. Ashcroft, 123 S. Ct. 769 (2003), filed May 20, 2002.

equal to the life of the author plus fifty years, raised to seventy years in 1998 by the Sonny Bono Copyright Term Extension Act. The 1976 Act fixed a term for works of hire[2] of seventy-five years from publication or 100 years from creation, whichever expired first; the Sonny Bono Act extended these terms to ninety-five and 120 years. The 1976 Act also made copyrights on works created after January 1, 1978, nonrenewable, but it allowed assignments and other transfers of copyrights to be terminated by the author or his heirs thirty-five years after the assignment or transfer.

The Constitution authorizes Congress to create copyright and patent protection only "for limited Times."[3] But the legal significance of "limited Times" is unclear, although the motivation—a hostility deeply rooted in Anglo-American law and politics to the conferral of monopolies by the executive branch of government[4]—is clear enough. Any time short of infinity, which is to say any fixed period of years, is "limited" in the literal sense of the word; and even if "limited" means something *far* short of infinity, this limitation might conceivably be got around by allowing repeated extensions of the copyright term. Renewals and extensions of patents and copyrights had been common in England in the eighteenth century, though on an individual rather than a wholesale basis, and it was English practice that provided the model and inspiration for the copyright clause of the Constitution and for the early federal copyright statutes.[5] And since common law copyright is perpetual, states could recognize copyright after the expiration of federal copyright protection if the federal copyright law disclaimed any intention of preempting state law.[6] Moreover, although Congress could not grant perpetual copyright under the authority of the Constitution's copyright clause, maybe it could do so under other grants of power to Congress, such as the power to regulate interstate and foreign commerce. That is unlikely; the framers clearly intended to limit as well as confer congressional authority to grant patents and copyrights. In any event, our concern is with the economics rather than the constitutionality of indefinite renewal.[7]

2. These are works in which the employer, or occasionally other hirers, of the actual author of the work owns the copyright unless the contract with the author provides otherwise. We examine the work for hire doctrine in Chapter 10.

3. On the significance of the term "limited Times," see Edward C. Walterscheid, *The Nature of the Intellectual Property Clause: A Study in Historial Perspective* 271–307 (2002).

4. See, for example, The Case of Monopolies, 11 Co. Rep. 85b, 77 Eng. Rep. 1260 (K.B. 1602).

5. See Walterscheid, note 2 above, at 355–356, 364.

6. See Bonito Boats, Inc. v. Thunder Craft Boats, Inc., 489 U.S. 141, 166–167 (1989).

7. The Supreme Court has recently upheld by a broad margin (7–2) the constitutionality of the Sonny Bono Act, see Eldred v. Ashcroft, 123 S. Ct. 769 (2003), against the claim that it violated the "limited Times" provision. In light of that decision, it is unlikely that a system of indefinite renewals, which has more to commend it than the Sonny Bono Act, would be held unconstitutional.

Although a copyright that could be renewed indefinitely might turn out to be perpetual, this would be unlikely for any but a tiny fraction of all copyrights. We shall see that fewer than 11 percent of the copyrights registered between 1883 and 1964 were renewed at the end of their twenty-eight-year term, even though the cost of renewal was small.[8] And only a tiny fraction of the books ever published are still in print; for example, of 10,027 books published in the United States in 1930, only 174 were still in print in 2001—1.7 percent.[9] These data suggest that most copyrights depreciate rapidly and therefore that few would be renewed if even a slight fee were charged; the sheer bother of applying for renewal appears to be a significant deterrent. Granted, it costs more to keep a book in print than to renew a copyright; and a copyrighted work's derivative works may have commercial value after the original work has lost it. Nevertheless it is apparent that even with an unlimited right of renewal the public domain would remain a vast repository of intellectual "property" (legally, nonproperty) available for use without charge both by consumers and as a source of free inputs into the creation of new intellectual property.

Allowing unlimited renewals might, depending on the length of the initial term and on the fee structure, actually *expand* the number of works in the public domain. The average value of those works would probably fall, since copyright in the most enduringly popular works would be renewed over and over again. But the total value might well rise, and not only because fewer works would remain under copyright for as long a time as under the present system. The public domain is not a fixed supply of works from which any enlargement of copyright protection subtracts. Its size is a positive function of the extent of copyright protection. The more extensive that protection is, the greater the incentive to create intellectual property, some fraction of which will become a part of the public domain when the copyright expires or, under the system we are suggesting, is not renewed. Cutting the other way, though, is the fact that a stiff renewal fee increases the expected cost of copyright and so may deter the creation of some expressive works.

8. The renewal fee was $1 from 1909 to 1947, $2 from 1948 to 1965, $4 from 1966 to 1977, $6 from 1978 to 1990, $12 from 1991 to 1992, $20 from 1993 to 1999, and $45 from 2000 to the present. Prior to 1992, a copyright holder who wanted to renew his copyright had to file a renewal application during the last year of the initial copyright term. An amendment that year to the Copyright Act made renewals automatic, although there still are some benefits to filing for renewal registration. See Robert A. Gorman and Jane C. Ginsburg, *Copyright: Cases and Materials* 356–357 (6th ed. 2002).

9. These data were computed from *American Library Annual and Book Trade Almanac for 1872–1957; The Bowker Annual* (same publication, new title) for 1974; and *Books in Print*, at *Bowker.com*.

The Benefits of Time-Limited Copyrights

Two propositions are widely believed by most economists; it is the tension between them that makes the question of a limited versus indefinite copyright term an interesting and difficult one. The first proposition is that, so far as is feasible, all valuable resources, including copyrightable works, should be owned, in order to create incentives for their efficient exploitation and to avoid overuse. The second proposition is that copyright should be limited in duration. The reasons offered in support of the second proposition should be familiar to the reader from the earlier chapters in this book: (1) tracing costs increase with the length of copyright protection; (2) transaction costs may be prohibitive if creators of new intellectual property must obtain licenses to use all the previous intellectual property they seek to incorporate; (3) because intellectual property is a public good, any positive price for its use will induce both consumers and creators of subsequent intellectual property to substitute inputs that cost society more to produce or are of lower quality, assuming (realistically however) that copyright holders cannot perfectly price discriminate; (4) because of discounting to present value, incentives to create intellectual property are not materially affected by cutting off intellectual property rights after many years, just as those incentives would not be materially affected if, during the limited copyright term, lucrative new markets for the copyrighted work, unforeseen when the work was created, emerged;[10] (5) in any event, retroactive extensions of copyright should not be granted. On the one hand, they can't affect the incentive to create new works, since a retroactive extension affects only the return on works already in existence.[11] On the other hand, the possibility of obtaining retroactive extensions invites rent seeking, as in the Disney Company's lobbying for the Sonny Bono Act, of which more shortly.

Determining the optimal term of copyright protection requires balancing at the margin the incentive effects of a longer term against the administrative and access costs, bearing in mind that the relevant access includes that of future creators of intellectual property as well as that of consumers of the existing property. Since, given discounting and depreciation, the incremental in-

10. One must be cautious, however, in asserting that "unforeseen" opportunities will not affect incentives. A particular new market may be unforeseen or unanticipated yet may be part of a *class* of markets that when the work was created had a foreseen, positive probability of coming into existence and therefore may have influenced the incentive to create the work. See generally Jane C. Ginsburg, "Copyright and Control over New Technologies of Dissemination," 101 *Columbia Law Review* 1613 (2001).

11. This is a slight overstatement. Knowledge of the possibility of a future lengthening of the copyright term might have some, though probably very small, incentive effects.

centive to create new works as a function of a longer term is likely to be very small beyond a term of twenty-five years or so,[12] administrative and access costs will tend to dominate, implying an optimal copyright term considerably shorter than the current term of life plus seventy years.[13] Thus the second proposition denies the first (valuable resources should be owned) and asserts that copyrightable intellectual property should be taken out of private ownership and placed in the public domain after a period of years no longer than required to generate socially efficient incentives to create new works. But is the second proposition always sound? It may be, since from a social standpoint (sometimes from a private one as well) property rights often cost more than they are worth. It has seemed so to most students of copyright law. But we have our doubts.

The argument for limiting copyright duration because of tracing costs is superficial except in explaining why *common law* copyright (largely extinguished, however, by the Copyright Act of 1976) in unpublished works is perpetual: there is usually only one copy of such works, so the cost of determining the copyright holder's identity is trivial unless the copy has passed through many hands.[14] Even in the case of published works, the costs of tracing the ownership of copyrighted works could be reduced to a low level by modest institutional reforms. It is true that enormous tracing costs would be incurred by any would-be publisher of a new translation of the *Iliad* if the heirs of Homer could enforce copyright in the work. But this is only because no one knows who they are. Equally great tracing costs would be required to determine the ownership of a parcel of land if land titles weren't recorded in

12. Suppose a copyright on a particular work would yield $1 per year in perpetuity at a discount rate of 10 percent. Under a system of perpetual copyright, the present value of this infinite stream of income would equal $10 ($= 1/r$). Under a limited copyright term ($= t$) the present value would be $(1 - e^{-rt})/r$. So if $t = 25$ and $r = .10$, the present value of $1 per year for twenty-five years is $9.18, which is more than 90 percent of the present value of a perpetual copyright. If the value of the copyright depreciates by, say, 5 percent per year, the difference in present value between a perpetual and twenty-five-year copyright is only about 2.5 percent ($6.67 versus $6.51).

13. For empirical evidence that the twenty-year extension of the copyright term by the Sonny Bono Act did not increase the production of copyrightable works, see Kai-Lung Hui and I. P. L. Png, "On the Supply of Creative Work: Evidence from the Movies," 92 *American Economic Review Papers and Proceedings* 217 (May 2002). See also Avishalom Tor and Dotan Oliar, "Incentives to Create under a 'Lifetime-Plus-Years' Copyright Duration: Lessons from a Behavioral Economic Analysis for *Eldred v. Ashcroft*," 36 *Loyola of Los Angeles Law Review* 437 (2002).

14. An example is the discovery of an unpublished manuscript in the library of Harvard University of a novel entitled *Inheritance*, written by Louisa May Alcott in 1849. It had been miscatalogued for many years and no one knew of its existence. Although Alcott was childless, the copyright holders—fourth-generation descendants of Alcott's father—were not difficult to locate. See Lawrence Van Gelder, "Uncovered at Harvard: Alcott's First Novel," *N.Y. Times*, May 1, 1996, p. C15.

public registries. It is not perpetual property rights but absence of registration that creates prohibitive tracing costs.

Were a system of indefinitely renewable copyright to be instituted today, there would be no great difficulty in identifying copyright owners a century or for that matter a millennium hence if, for example, the law required copyright owners to reregister their copyrights every ten or even twenty-five years in some central registry under the name of the copyright holder and to notify the registry in the event the copyright was transferred. (Owners of existing copyrights would be required to register them upon the creation of the new system.) The owner would be required to provide the registry with his address and notify it of any changes of address; a transferee would likewise be required to furnish this information to the registry. Then a search of the registry either under the name of the owner or under the title of the copyrighted work (as the work itself might not reveal the name of the original copyright owner) would reveal the address of the current copyright holder, his agent, etc., from whom a license would have to be sought, just as in the case of the registries in which titles to real estate are recorded and the Uniform Commercial Code registries in which security interests in personal property are recorded. A fee would be charged for renewing a copyright registration to recover the costs imposed on the registry itself and on the searchers by the renewal. The fee could exceed those costs if it were desired to expand the public domain by discouraging renewals of works unlikely to have much commercial value.

Under existing law, when copyright protection begins is relatively unimportant because the duration of protection is determined not by that starting point but instead (except in the case of works for hire) by the death of the author. Under a system of renewals, the starting point becomes critical. So our suggested system would require a return to something like the pre-1976 law, when copyright protection generally began with publication or registration.[15]

There would be no need to require, in addition to registration, that a notice of copyright be placed on copyrighted works indicating the name of the copyright holder and the date of the most recent copyright registration or renewal, as under the 1909 copyright act. The registry should provide adequate notice. And for some works up-to-date notices are either infeasible or simply too costly in relation to the benefits in reducing tracing costs. For example, the seller of a work of art who retains the copyright cannot place an up-to-date copyright notice on the work when it has been out of his possession for many years. There is also a risk of cluttering up a work of art with multiple notices that would detract from the work's artistic merits, unless the

15. This would require the United States to withdraw from the Berne Convention, of which it became a signatory in 1989 and which requires signatories to provide a minimum term of life plus fifty years.

notice had to be placed only on the back of the canvas, or the back of the pedestal on which a sculpture stands.

Joint ownership of copyrights is possible, but we need not worry that the more owners there are, the greater the tracing costs. Since any joint owner of a copyright may license its use subject only to a duty to account for the profits to the other owners, a would-be licensee of a jointly owned copyright need find only one of the joint owners to negotiate with.

The transaction-costs argument against indefinite renewal is stronger than the argument from tracing costs but must not be exaggerated either. Although transaction costs would be incurred each time a copyright was renewed, consisting mainly of the time costs of the copyright holder and the costs of administering the renewal system, they would be slight if most copyrights were not renewed—and the longer the initial term and the higher the renewal fee, the fewer would be renewed. However, the costs incurred in negotiating for the licensing of one or more of that minority of works on which the copyright would be renewed many times would be higher than under the present system, though not when the new work would be copying only one old work—for example, Joyce's *Ulysses* and Homer's *Odyssey,* the movie *Clueless* and the novel *Emma, My Fair Lady* and *Pygmalion, West Side Story* and *Romeo and Juliet, Ragtime* and *Michael Kohlhaas.* But sometimes multiple works are copied, as in Manet's *Déjeuner sur l'herbe* and T. S. Eliot's *The Waste Land,* both celebrated works that would be likely to remain under copyright indefinitely under a system of indefinite renewals. However, a work that borrows from multiple works is unlikely to offer itself as a substitute for any of them, especially when the borrowing from each one is small, and so such a work should probably be shielded from liability in any event by the fair use doctrine, which we argue repeatedly in this book—most pertinently to the present issue in Chapter 9—should be construed generously.[16] Furthermore, the degree to which even a copy of a single work should be deemed infringing depends on the interpretation given to "substantial similarity," the criterion of infringement as we noted in Chapter 4. Although *Ulysses* is so heavily indebted to the *Odyssey* that it could be regarded as a burlesque of the earlier work, the debts are so well concealed from any but the most learned reader that probably the two works would not be deemed substantially similar. A narrow construal of substantial similarity and a broad construal of fair use would help maintain an ample public domain under a system of indefinite renewals.

16. Consider the following hypothetical example offered in opposition to the Sonny Bono Act: "While interviewing students for a documentary about inner-city schools, a filmmaker accidentally captures a television playing in the background, in which you can just make out three seconds of an episode of 'The Little Rascals.' *He can't include the interview in his film unless he gets permission from the copyright holder to use the three seconds of TV footage.* After dozens of phone calls to The Hal Roach Studios, he is passed along to a company lawyer who tells him that he can include the fleeting glimpse of Alfalfa in his nonprofit film, but only if he's willing to pay $25,000. He can't, and so he cuts the entire scene." Jeffrey Rosen, "Mouse Trap: Disney's Copyright Conquest," *New Republic,* Oct. 28, 2002, p. 12 (emphasis added). We have italicized the sentence that indicates a much too narrow understanding of fair use. The "three seconds" quotation is an example, merely visual rather than verbal, of the kind of brief quotation from copyrighted works that the fair use doctrine, for economic reasons explained in Chapter 4, permits without need for a license from the owner of the copyright. Such misunderstandings are widespread. One of us has encountered an academic publisher who insisted that any quotation of two or more lines of copyrighted poetry requires a license to reprint. That is unsound as a matter of both economics and law.

Transaction costs under such a system would be greatest for explicitly composite works, such as anthologies of well-known earlier works. Under existing law the publisher of a collection of the world's greatest poems need obtain copyright licenses for only a subset of the poems—none first published before 1923.[17] Under a regime of indefinitely renewable copyright instituted in 1500 A.D., most of the poems in an anthology of popular poetry might still be under copyright and therefore many more licenses would have to be obtained for a new anthology.

The aggregate transaction costs of a system of indefinite copyright renewals would depend on the number and possibly the value of licenses (holding tracing costs constant), the transaction costs per license, and the administrative cost of operating the renewal system. Since the number of licenses would depend in part on the total number of works renewed, aggregate transaction costs could actually fall compared either to a system of automatic renewals or to a single term of life plus seventy years.

Consider now the public-good argument for limiting the duration of a copyright. Copyright law permits a writer or publisher or other producer of an expressive work to charge a price that exceeds marginal cost, which may be very low. To extend the copyright term is to increase the duration of the restriction on output and likewise the producer's revenue. As there is no solid basis for thinking the existing term too short to enable producers of expressive works to cover their full costs, there is no good incentive-based argument for the extension. Even if one thought the current level of copyright protection inadequate, extending the term, as by making copyrights renewable indefinitely, would do little to correct the balance, given discounting to present value. It might seem, therefore, that the only effect of extending the copyright term would be to create access costs by reducing the number of works at any given time that are in the public domain and can therefore be used without need to obtain a license.

17. The copyright on a work first published in 1922 would terminate after seventy-five years (a twenty-eight-year initial term plus a forty-seven-year renewal term), and thus in 1997. The Sonny Bono Act adds an additional twenty years of protection for works that were still under copyright in 1998.

But those costs are unlikely to be great. First, just as future revenues must be discounted to present value to determine their value, so future costs must be discounted similarly to determine their present cost. If the present value of some remote future benefit is trivial, so is the present cost of the equally remote future deadweight loss. It is true that when values cannot be monetized the use of a social discount rate, somehow computed, may be preferable to the use of the private discount rate. But deadweight costs are readily monetizable. Given an estimate of their dollar cost in some future period, government can offset them by investing a sum equal to their present value, computed according to current long-term interest rates, in financial instruments.

The discounting of future deadweight costs ceases to reassure if the question is whether to extend the term of existing rather than future copyrights, at least existing copyrights that, unless extended, will soon expire. Suppose a copyright that was about to expire is extended another twenty years. The deadweight costs will begin to accrue immediately. They still must be discounted, but the present cost will be much greater than if the discounting were of deadweight costs to be incurred in a period first beginning seventy-five years from now.[18] The case for a system of indefinite renewals may thus be stronger if it is limited to copyrights obtained after the system is instituted. However, a potentially offsetting benefit of a system of indefinite renewals not limited to future copyrights—reduced rent seeking—will be noted shortly.

Second, because the scope of copyright protection is narrow, the size of the deadweight loss created by copyright protection will usually be small. The narrow scope of the property right implies the existence of close substitutes, which increase the elasticity of demand for the copyrighted work. In the simple case of a linear demand curve and constant marginal cost, the deadweight loss of monopoly is one-half the amount by which the monopolist's revenue exceeds his cost; the higher the elasticity of demand, the smaller that amount. It has been argued that the optimal duration of a patent would be infinite if the scope of patent protection were narrowed appropriately.[19] Conceivably the scope of copyright protection is already so narrow that an infinite copyright term would not be a source of significant deadweight loss. But this is merely a conjecture (are there good substitutes, for example, for Shakespeare's plays or Mozart's piano concertos?), and one reason the scope is narrow is that the public domain provides a source of free inputs into the

18. This point was emphasized in the economists' amicus curiae brief in the *Eldred* case, note 1 above, at 11.

19. Richard Gilbert and Carl Shapiro, "Optimal Patent Length and Breadth," 21 *RAND Journal of Economics* 106 (1990).

creation of new copyrightable works. If valuable works are withheld from the public domain because the copyright term has been extended, there may be significantly fewer public domain works (weighting number by quality) upon which to draw, which will reduce competition with existing copyrightable works.

Third, indefinite renewal and extending the copyright term are not the same thing. The length of the initial and renewal terms, the fee charged for renewal, and the scope of renewal (would it be limited to a single work, or could it cover a group of works?) are variables that can be adjusted to produce on average whatever de facto copyright term is deemed socially desirable; nor need the length, fee, or scope be the same for all classes of work (books, software, music, etc.). The shorter the initial grant (it could be as short as ten years) and the higher the renewal fee, the shorter would be the de facto copyright term of most works (all, if the renewal fee is high enough—and it could escalate with the number of renewals) and so the fewer the number of works that would be protected by copyright. It is the *composition* of the public domain that would most likely be changed by an indefinite-renewal system because there would be better sorting of works into two categories: (1) valuable works, where the benefits of property rights may well exceed the costs, and (2) works of little value, where the costs of administering copyright protection are very likely to exceed the benefits and a stiff renewal fee would discourage the owner from seeking continuing copyright protection. The public domain would be enlarged, although some valuable works would be excluded that under the present system fall into the public domain eventually.

But the following complication would attend stiff renewal fees. When the fees are very low, the fact that the commercial value of copyrightable works varies enormously is relatively unimportant, except perhaps for photographs, since each photograph is a copyrightable work and serious photographers take many pictures in a single session. But when renewal fees are high, uniform fees can become prohibitive for many copyrighted works—not only photographs. It would be one thing to charge $1,000 to renew the copyright on a movie after ten years, and another to charge $1,000 to renew the copyright on an academic book. But while group renewals of photographs would have to be allowed (as they are under current law), the fact that stiff fees would deter the renewal of copyrights on works having little commercial value is not necessarily a bad thing. By definition these are works unlikely to yield the owners of the copyrights on them a significant return, so it is best that they be placed in the public domain where they can be used as free inputs into the production of new intellectual property, though we shall have to qualify this point later. Still, a single fee for all types of copyrighted work is

unlikely to be optimal. An alternative that would minimize legislative and regulatory discretion, and hence rent seeking, would be to make the fee equal to a fixed percentage of the first year's inflation-adjusted revenues from the sale or rental of the copyrighted work.

We have a related response to the argument that under a system of renewals many copyrights would be lost by sheer ignorance or inadvertence on the part of copyright holders. One of the reasons given for the abandonment of the renewal system by the Copyright Act of 1976 was that many copyright holders didn't know they had a renewal right or made fatal errors in filling out the renewal form. (The problem was solved by switching to a system of life plus years, since no one forgets to die.) But ignorance is endogenous; care and information are costly goods, which are not "bought" if the benefits from them are slight. It was because so many copyrights had (and have) so little value that so many copyright holders were not assiduous in protecting their copyrights.[20]

With regard to rent-seeking implications of a switch to a system of indefinite renewals, the first thing to note is that owners of copyrights on old but still commercially valuable works have an incentive to expend resources on lobbying Congress to extend the copyright term on these works. Retroactive extensions do not enhance incentives to create expressive works, so if those incentives are the only benefits from copyright, such extensions will increase transaction and access costs without generating any offsetting value. The costs of Disney's successful efforts to lobby for the Sonny Bono Act that retroactively extended its copyrights by twenty years in order to protect its soon-to-expire copyrights on Mickey Mouse and other cartoon characters,[21] and the costs of the unsuccessful efforts (which seem, however, to have been slight) of competing interests to oppose the extension, were incurred to obtain and limit economic rents, respectively. If there are no offsetting social benefits from retroactive extensions of the copyright term, these costs were wasted from a social standpoint.

20. Granted, the renewal provision of the 1909 Copyright Act was unclear with respect to the prospective assignability of the renewal term and to the respective rights of an assignee and a statutory successor (the provision authorized renewal by the author's heir if the author died before the initial term expired). See Pierre N. Leval and Lewis Liman, "Are Copyrights for Authors or Their Children?" 39 *Journal of the Copyright Society* 1 (1991). These uncertainties, which any law reinstituting renewal could easily dispel, may have been responsible for some failures of renewal.

21. See Janet Wasko, *Understanding Disney: The Manufacture of Fantasy* 85–86 (2001). The Act went into effect on October 27, 1998, four years before the copyright on the original Mickey Mouse would have expired. The Mickey Mouse character has been altered in appearance several times, however, as we'll note in a moment, and the successive versions (which of course are derivative works) copyrighted, so even if the original copyright had been allowed to expire, Disney could have fended off efforts to market an exact duplicate of the current Mickey Mouse.

Rent-seeking activities are a natural consequence of any fixed copyright term, since the Congress that enacts the term cannot prevent future Congresses from increasing it retroactively. There will always be copyright holders whose incomes are diminished when their works fall into the public domain, and they have an incentive to expend resources on seeking retroactive extensions as the end of the copyright term draws near. This rent seeking would be lessened by indefinite renewability, which by eliminating the prospect of losing the income produced by old but still valuable copyrights would largely eliminate the incentive to lobby for copyright extensions. "Largely" rather than "entirely" because some resources might still be expended on lobbying for lower renewal fees and longer renewal terms. But normally it would be cheaper to pay the renewal fee than to try to change the law. And while a system of indefinite renewals limited to future copyrights would not curb the incentive to seek retroactive extensions of existing copyrights, that problem would disappear in time as old copyrights lost their value.

A particularly unfortunate aspect of congressional extensions of the copyright term is that they apply to *all* copyrighted works, not merely those whose owners lobbied for the extensions. Disney might have sought a private bill that would have extended only its copyrights, but a politically more effective strategy was to ally itself with other copyright owners and represent that the legislative relief sought would benefit all creators of intellectual property, not just Disney and a handful of rich heirs. But the result of the blanket extension of the copyright term was that a huge amount of intellectual property having little or no commercial value, yet potential value as a public domain input into future intellectual property, will be kept out of the public domain for another twenty years.

It might be objected that allowing indefinite renewals would eliminate only one form of rent seeking because copyrights have other dimensions of value besides duration, notably scope. But whatever incentive there is for lobbying for enlarged scope exists under the current system; it would not be greater under a system of indefinite renewals. A more serious concern is that copyright holders might renew their copyrights for strategic purposes, hoping one day to "hold up" an author who wanted to copy their work. This practice would resemble strategic patenting, discussed in Chapter 11, and would be a danger posed particularly by software copyrights, which is one reason why we do not discuss the applicability of a system of indefinite renewals to them. With regard to other classes of copyrightable work, a stiff renewal fee, combined with the effect of discounting to present value, should minimize the problem.

If we are correct so far, the *average* copyright duration might be shorter under a system of indefinite renewals than under the current system. Such a system might therefore reduce access costs for most but not all works, com-

pared to the present system. (It probably would not reduce deadweight costs because these presumably are generated mainly by valuable copyrights, which would tend to be renewed.) But whatever the case with regard to costs, it may seem that there could be no social benefits from continuing indefinitely to protect even a small number of valuable works. This assumes, however, that the only justification for copyright protection is the incentives it produces to create new works. This may be wrong, as we consider next.

The Social Benefits of Allowing Some Copyrights to Remain in Effect Indefinitely

We focus here on the relative handful of copyrights that under a system of indefinite renewals could be expected to remain in force for even longer than life plus seventy years. Because of discounting to present value, an increase in the copyright term could not be justified on the ground that it would enhance the incentives to create expressive works. But the economic theory of property rights emphasizes not only their incentive effects, that is, the investment they encourage, but also their effect in optimizing current uses of property. We recall from Chapter 1 that because a natural pasture is not created by human effort, there is no social value in encouraging investments in creating it, but that in the absence of property rights the pasture would be overgrazed; none of the users would take account of the cost that his use imposed on the other users by making their cattle graze more to obtain the same amount of food, and thus gain less weight. Moreover, not all investments in expression are made before copies of a work are sold; some, which we'll call "maintenance" investments, are made afterwards and may be discouraged by durational limits on copyright. So in this part of the chapter we shall be considering congestion externalities and maintenance incentives as arguments for making copyrights indefinitely renewable.

CONGESTION EXTERNALITIES Benefits of property rights analogous to those of property rights in natural pastures have already been recognized in two areas of intellectual property law. One is trademark law, which, as we know, does not impose any fixed limitation on the duration of a trademark, since confusion would result if the same trademark denoted goods of different provenance and quality, and which, through the concept of blurring (a subcategory of dilution), protects trademark owners from the loss of value resulting from nonconfusing duplication of their trademarks, a form of overuse. The other area is the law of publicity rights, which as we noted briefly in Chapter 2 prevents the use of one's name or likeness in advertising or for other commercial purposes without one's permission. The tendency is to

make publicity rights inheritable.[22] The motive is not to encourage greater investment in becoming a celebrity (the incremental encouragement would doubtless be minimal),[23] but to prevent the premature exhaustion of the commercial value of the celebrity's name or likeness.[24] The analogy to overgrazing is close. Overgrazing causes crowding in the short run, with a resulting reduction in weight gain, and depletion of the pasture in the long run with similar, though possibly more drastic, results. Similarly, overexposure of a celebrity may turn people off in the short run and truncate the period in which his name or likeness retains commercial value.

Recognition of an "overgrazing" problem in copyrightable works has lagged. Typical is the statement endorsed by many professors of intellectual property law in opposition to the Sonny Bono Act:

> The fundamental difference between tangible and intellectual property is that intellectual property is a nondepletable commons, while tangible property necessarily depletes with use. "The tragedy of the commons" is that failure to recognize perpetual and transferable property rights in tangible property leads inevitably to "overgrazing," as soon as an item of property enters the public domain from which everyone may draw freely. Recognition of perpetual property rights leads to economic efficiency, because a rational owner will optimize the balance between present and future consumption.
>
> There can be no overgrazing of intellectual property, however, because intellectual property is not destroyed or even diminished by consumption. Once a work is created, its intellectual content is infinitely multipliable.[25]

This is overstated, if only because it ignores the trademark and right-of-publicity cases that recognize that intellectual property *can* be diminished by

22. See Huw Beverley-Smith, *The Commercial Appropriation of Personality* 184 (2002); Mark F. Grady, "A Positive Economic Theory of the Right of Publicity," 1 *UCLA Entertainment Law Review* 97, 124–126 (1994). See also Douglas G. Baird, "Does Bogart Get Scale? Rights of Publicity in the Digital Age," 4 *Green Bag* (2d ser.) 357, 363–364 (2001).

23. As argued in Michael Madow, "Private Ownership of Public Image: Popular Culture and Publicity Rights," 81 *California Law Review* 125, 205–215 (1993). For an exception, see Zacchini v. Scripps-Howard Broadcasting Co., 433 U.S. 562 (1977), discussed in Madow, above, at 208 n. 395.

24. See Grady, note 22 above, at 103, 126; Richard A. Posner, *The Economics of Justice* 248 (1981). The English call this "face wearout." Madow, note 23 above, at 222. Madow, however, is skeptical about its empirical significance. See id. at 221–225.

25. Denis S. Karjala, "Statement of Copyright and Intellectual Property Law Professors in Opposition to H.R. 604, H.R. 2589, and S. 505, The Copyright Term Extension Act, Submitted to the Joint Committees of the Judiciary," Jan. 28, 1998, *http://www.public.asu.edu/~dkarjala/legmats/1998statement/html.*

consumption. But assessment of the welfare effects of congestion requires distinguishing technological from merely pecuniary externalities (transfer payments). The externality in the pasture case is technological because it imposes a real cost (diminished weight gain) rather than merely altering the distribution of wealth. Refusing to recognize inheritable publicity rights could impose either type of externality or both types. If anyone could use Humphrey Bogart's name or likeness in advertising, the aggregate value of that advertising use might be greater even though Bogart's estate would lose income. Indeed, if the marginal cost of additional copies of his image were zero, the marginal utility would also be zero, even though the total utility could be very great. But the total utility might decline if the lack of excludability and resulting proliferation of the Bogart image led to confusion, the tarnishing of the image, or sheer boredom on the part of the consuming public. Eventually the image might become worthless.

Could this be a problem with regard to copyrightable expression? There is some evidence that it is a concern of the Walt Disney Company with regard to its copyrighted characters, such as Mickey Mouse. "To avoid overkill, Disney manages its character portfolio with care. It has hundreds of characters on its books, many of them just waiting to be called out of retirement . . . Disney practices good husbandry of its characters and extends the life of its brands by not overexposing them . . . They avoid debasing the currency."[26]

Figure 8.1 illustrates the problem. D^0D^0 is the demand schedule in period t for some expressive work. Obviously copyright protection in period t and all future periods would have no effect on whether the work was created in period $t = 0$, but it would create a deadweight cost, illustrated by the triangle $P^0Q^2Q^0$, as a result of the copyright holder's charging P^0 when his marginal cost is zero. Terminating the copyright in t would eliminate the deadweight loss, as the number of uses of the work would increase to Q^2, that is, to the point at which the value of the marginal use equaled zero. But now suppose that contrary to the usual assumption about copyrights, additional uses impose technological externalities. Then terminating the copyright will lead not only to a movement along the demand curve but also to a downward shift (say to D^0D^1) in the overall demand, destroying value equal to the difference between the area under the original demand curve D^0D^0 up to P^0 and the area under D^0D^1 up to a zero price. If the externalities are small, the difference between the two demand curves may be negative, so that terminating the copyright at t would increase value. But if they are large, termination can result in a net loss in value. In the limit, additional uses beyond Q^0 might depress the demand curve (as it rotates downward around the point that intersects the vertical axis at D^0) until it coincided with the vertical axis. In that event, terminating the copyright would destroy all its value in period t—the area under D^0D^0 between the point at which it intersects the vertical axis and P^0—and presumably in all future periods as well.

26. Bill Britt, "International Marketing: Disney's Global Goals," *Marketing*, May 17, 1990, pp. 22, 26.

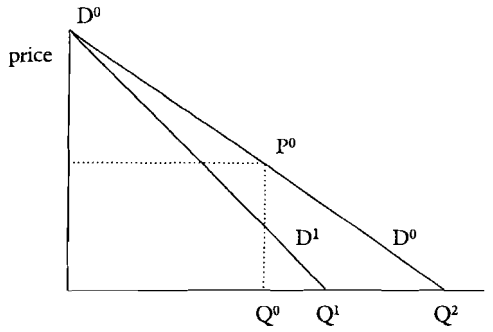

Figure 8.1. Copyright externalities

A book or other copyrightable property is conventionally regarded as a public good in the sense that unlike a pasture its use by one consumer does not interfere with its use by any other. This point cannot be decisive, however; a celebrity's name or likeness is a public good in the same sense, yet unlimited reproduction of the name or the likeness could prematurely exhaust the celebrity's commercial value, just as unlimited drilling from a common pool of oil or gas would deplete the pool prematurely. The same could be true of a novel or a movie or a comic-book character or a piece of music or a painting, particularly with regard to copyrights on components of completed works rather than on the completed works themselves. That is why it is better to say that a book or other copyrighted work has public-good characteristics than that it *is* a public good.

If because copyright had expired anyone were free to incorporate the Mickey Mouse character in a book, movie, song, etc., the value of the character might plummet. If this came about only as a result of a movement down the original demand curve, the ordinary consequence of an increase in output, aggregate value would actually increase, by the area under the demand curve between P^0 and a zero price. Alternatively, however, the public might rapidly tire of Mickey Mouse; and his image might also be blurred or even tarnished, as some authors portrayed him as a Casanova, others as catmeat, others as an animal-rights advocate, still others as the henpecked husband of

Minnie. There would thus be both a movement along and a downward shift of the demand curve in Figure 8.1 until Mickey Mouse's commercial value was zero. The same thing would happen in the absence of negative technological externalities, but since the demand curve would be unchanged, total value would increase because there would be no deadweight loss.

To the extent that unauthorized reproductions of Mickey Mouse were classified as parodies, they would be immunized from liability by the fair use doctrine even in a regime of indefinitely renewable copyright. But not all would be parodic. The "Here's Johnny" right-of-publicity case mentioned in Chapter 6 is a good illustration of how an image can be degraded by the context in which it is presented to the public. The "character" in that case was a real human being, not a fiction, but analytically there is no difference between the congestion problem faced by a celebrity and that encountered by a fictitious character. If, as economic analysis of the right of publicity suggests, there is a real congestion problem in the former case, there must be in the latter case as well.

This analysis gains support from the discussion of derivative rights in Chapter 4. One purpose of giving the owner of a copyright a monopoly of derivative works is to facilitate the scope and timing of the exploitation of the copyrighted work—to avoid, as it were, the "congestion" that would result if once the work was published anyone could make and sell translations, abridgments, burlesques, sequels, versions in other media from that of the original (for example, a movie version of a book), or other variants without the copyright owner's authorization. The result could be premature saturation of the market, consumer confusion (for example, as to the source of the derivative works), and impaired demand for the original work because of the poor quality of some of the unauthorized derivative works.

We must not press the congestion argument (or its guilt-by-association cousin, illustrated by the "Here's Johnny" case and by some of the trademark cases discussed in Chapters 6 and 7) too far.[27] While examples can be given of works even of elite culture that may have been damaged by unlimited reproduction (the *Mona Lisa,* the opening of Beethoven's *Fifth Symphony,* and several of Van Gogh's most popular paintings come immediately to mind), there are counterexamples: the works of Shakespeare seem unimpaired by the uncontrolled proliferation of performances and derivative works, some of them kitsch, such as Shakespeare T-shirts and the movie *Shakespeare in Love.* And

27. A technical legal argument for not doing so is that the patent and copyright clause of the Constitution authorizes the grant of patents and copyrights "to promote the Progress of Science and useful Arts," and the elimination of congestion externalities might seem remote from that goal, though we think not, since such externalities impede the efficient functioning of the "Science" and "useful Arts" markets, that is, the markets in intellectual property.

in the field of popular culture, think only of Santa Claus as an example of the power of an iconic character to survive incessant use, apparently undamaged—but with the difference that he is strictly a seasonal character; we have eleven months of respite from him. Much of the Disney Company's own considerable commercial success has been based on its use of fictional characters that are in the public domain, such as Pinocchio, Cinderella, and Quasimodo. Still, the right-of-publicity cases show that there is potentially a legitimate concern here, one that economic analysis should not ignore completely. Think of the role of prominent "literary widows," such as Valerie Eliot and Sonia Orwell, in managing the copyrights of their husbands—would there not have been a danger, had Orwell's copyrights expired on or shortly after his death, of an avalanche of Orwell derivative works that might have made the world tire of him?

In a later chapter we shall express skepticism concerning moral rights doctrine. Among other things, the doctrine confers on creators of intellectual property, mainly artists, the right to prevent their work from being disfigured, even if they have assigned their copyright. We may seem inconsistent in suggesting that copyright owners be entitled to prevent the overuse of their work. There is no inconsistency. Artists can control the use of their work by retaining copyright, since, as we know, the copyright owner has the exclusive right to authorize derivative works: a Mickey Mouse who speaks foul language is a derivative work of the copyrighted Mickey Mouse and so the Disney Company can prevent its creation unless the creator can shelter under the fair use umbrella for parodies. It does not follow that if Disney assigned its copyright, it would still have a right, as under moral rights doctrine (were it available to corporations), to prevent the creation of a degrading version of Mickey Mouse. What is at issue, moreover, in our discussion of congestion externalities is not artistic integrity or reputation, values that may have noncommercial aspects, but a purely economic concern with minimizing economically harmful externalities.

The concern we have expressed regarding congestion externalities may seem to argue for a greatly enlarged concept of trademark dilution, a doctrine that we discussed critically at the end of the preceding chapter. Just as the promiscuous use of Humphrey Bogart's name and likeness in advertising might reduce the aggregate value of that advertising use, so might the promiscuous use of the name Rolls Royce as a trademark reduce the aggregate advertising value of the name even if the consuming public was not confused about the identity of the users. But probably antidilution law is not needed to prevent this problem from arising. The reason is that any widespread use of the mark *would* be confusing. Suppose that a hotel chain adopted the name Rolls Royce. Knowing that modern corporations are often highly diversified

and that both automobiles and hotels are part of the travel sector of the economy, many consumers would infer an affiliation between the automobile company and the hotel chain. At the other end of the economic ladder, if, impressed by the success of a Rolls Royce hot-dog vendor, a vendor of roast chestnuts adopted the name Rolls Royce, there would again be a danger of consumer confusion—confusion not with the automobile manufacturer but with the hot-dog vendor; people might think the hot-dog and roast-chestnut vendors affiliates. If, however, somehow without confusion as to source, a prestigious name became so widespread as to threaten supersaturation and resulting loss of value, and thus create a true congestion externality, there would be an argument for invoking the antidilution principle.

MAINTENANCE INCENTIVES The conventional economic criticism of the length of the copyright term draws too sharp a distinction between creation and copying. Imagine a novel published many years ago in which copyright has expired. The novelist is rediscovered and there is a surge in demand for his novels. Since no publisher could establish a property right in them, the incentives of publishers to publish and promote them might well be inadequate from a social standpoint. Often the demand for particular works of intellectual property is unknown before they actually hit the market.[28] Suppose an enterprising publisher has only a 20 percent chance of success with obscure public domain authors. He publishes the works of five such authors in order to have one success. In the absence of copyright protection, other publishers can wait and see which author sells and then bring out their own versions of his works. Publishers who wait avoid the costs of failure, but their free riding on the market information developed by the first to publish reduces the incentive of any publisher to search for potentially successful public domain works.[29] The tendency would be for only works of already well-known and safe authors whose works were in the public domain to be published.

The gravity of this problem should not be exaggerated. There is plenty of publication of public domain works; we remind the reader of the discussion

28. This has been noted in connection with dolls and other goods portraying dead celebrities. See Ronald Alsop, "Items Portraying Dead Stars Produce Profits Controversy," *Wall Street Journal,* May 10, 1984, p. A37.

29. In other words, "a work's public domain status is far from an unqualified incentive for utilizing it . . . Some of the obvious concerns are whether a copyrighted derivative work will have to compete with other, often low-budget, low quality copies and whether the producer of the copyrighted derivative of a public domain work is likely to have anything unique in the long run." Arthur R. Miller, "Copyright Term Extension: Boon for American Creators and the American Economy," 45 *Journal of the Copyright Law Society* 319, 324 (1998).

in Chapter 2 of the incentives for publication that are independent of intellectual property law. But, grave or trivial, a system of indefinite renewals would ameliorate the problem. It would not solve it, however, since works that were no longer popular would tend not to be renewed many times. Indeed, the problem would exist even if copyright were perpetual. Owners of a perpetual copyright that they considered worthless would not take even modest steps to assure the continued registration (for example, notifying the registry of changes of address) that a system of perpetual, as of indefinitely renewable, copyrights would require be established. A complete solution would require that the saviors of old works on which copyright had expired without renewal, like finders in the law of real property, be allowed to obtain copyright in those works. We consider that possibility later. For now, it is enough to observe that a system of indefinite renewals would, depending on the renewal fee, on whether group renewals were permitted, and on the formalities involved in renewal, somewhat improve the incentives to invest in public domain works.

Under the present regime a publisher has an incentive to make changes in any public domain work that he does revive, since he can copyright the changes. But changes made merely to stake a claim, just like premature introduction of a new product in order to sew up a desired trademark, are inefficient. Extending the copyright term might thus reduce socially excessive product differentiation ("overmaintenance" as it were).

Conversely, if because of its age a newly resuscitated novel were in need of an elaborate scholarly apparatus, re-editing, or other costly additions to make it accessible to a modern readership, publishers might be reluctant to undertake the needed measures, even if they could copyright the scholarly apparatus (which they could not do to the extent it was deemed a matter of ideas rather than expression). They would fear that the cost could not be recouped in the face of competition from cheap, bare-bones editions of the novel. Reviewers might use the scholarly apparatus, of course without paying anything (not even the price of the book, since reviewers generally receive free review copies), to explain the book to the public, who would then buy the bare-bones edition.

Or consider an old movie on which copyright had expired that a studio wanted to issue in a colorized version that would be very expensive to prepare. Promoting the colorized version might increase the demand for the black and white version, a close substitute. Since anyone could copy and sell that version, the studio would have to take into account, in deciding whether to colorize, the increase in demand for the black and white version. As a result the expected revenue from colorization might be less than the private costs and so the studio would decide against it. Indefinite renewal might pro-

vide a complete solution, since, given the public's avidity for movies old as well as new, an old movie would be quite likely to have retained enough value to have warranted the expense of renewal.

We conjecture that the reason so few classical composers are recorded and performed is that it is more costly to produce a musical composition than it is, say, to photograph a painting. The recording company that discovered and revived the works of a forgotten or obscure composer would be risking a substantial amount of money in an uncertain venture that could be imitated if successful. Much less expense would be involved in publishing a book or even arranging an exhibition of works of a forgotten or obscure painter. The absence of property rights in the music of well-known classical composers may also explain why many different recording companies record the same public domain works of Beethoven, Mozart, Bach, and other well-known composers.[30] Recording companies differentiate their product by promoting the performer or artist who has signed an exclusive contract with the company. Because a recording company can, for example, copyright the Chicago Symphony Orchestra's recording of Mahler's *First Symphony,* it has an incentive to promote that version; it has little incentive to promote the public domain work of an unknown composer, since it could not appropriate the benefits of its promotional efforts, as distinct from benefits that might accrue from a recorded performance of the unknown composer's work by a popular performer.

Consider also the effect on the recording of a composer's obscure works when his copyrights expire. Our analysis implies that upon the expiration of Puccini's copyrights, the rate at which his obscure works were recorded fell relative to recordings of the best-known works, since an investment in creating a demand for the obscure works would be more difficult to recoup once the works were no longer under copyright.

These examples, unlike those used to illustrate the economic analysis of publicity rights and our extension of that analysis to copyright, show that a case against a definite time limit for copyrights can be grounded in the traditional incentive-based argument for property rights, though with a new twist. The new twist is recognition that the need to invest in intellectual property to maximize its value is not necessarily exhausted in the initial creation of the property. Investment may be necessary to maintain the value of the property and also to resurrect abandoned or otherwise unexploited intellectual property.[31] Magnitudes are critical, however, and in their absence only tentative conclusions are possible.[32] We are told that the Disney Corporation has spent tens of millions of dollars refurbishing the Mickey Mouse character, both by subtle alterations in the character and by situating it in carefully selected entertainment contexts in an effort to increase the appeal of Mickey Mouse to the current generation of young children. The incentive to make such expenditures would be impaired if the copyright expired, allowing anyone to use the character, though the copier could not copy any newly copyrightable features that Disney had added to the original character.[33] This would be an important qualification should it turn out that only the most recent version of the character retains commercial appeal, but that seems unlikely.

If this analysis is correct, the drumbeat of criticisms of the retroactive extension of the Mickey Mouse copyright overlooks valid, although in the present state of knowledge inconclusive, economic arguments for extended copyright protection that are independent of whether the protection is extended

30. Eight composers (Mozart, Beethoven, Bach, Brahms, Tchaikovsky, Schubert, Chopin, and Haydn) account for 47 percent of all classical recordings. See F. M. Scherer, "The Innovation Lottery," in *Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society* 3, 14 (Rochelle Cooper Dreyfuss, Diane Leenheer Zimmerman, and Harry First eds. 2001).

31. "For a work to be commercially successful, it requires effort and investment which, while not 'creative,' is still necessary to generate value. For example, authors employ literary agents, publishers advertise, etc. With musical composition and photographs, the collection, arrangement and indexing of the works adds value. With film, preservation requires constant attention. Even the straightforward act of printing a book entails a risk on investment. Arguably, none of these activities will be pursued as vigorously on behalf of public domain works as they are for works with ownership rights. And, from an economic point of view, these activities 'create' real value." Edward B. Rappaport, "Copyright Term Extension: Estimating the Economic Values" 4 (Congressional Research Service, May 11, 1998).

32. Rappaport states that the effect he describes "may be important in some cases, but, we believe, will more often be marginal." Id. He does not explain the reasoning or evidence that led him to this conclusion.

33. The public legislative history of the Sonny Bono Act contains little discussion of the twenty-year addition to the duration of copyrights on works for hire, such as the Disney copyrights. The main reasons given for the extension for individuals (remember that the Act extended the term for individuals from life plus fifty years to life plus seventy years) were (1) balance of payments and (2) the fact that people are living longer (Irving Berlin being a pertinent example). The second point makes very little sense, since an increase in longevity will automatically increase the length of the copyright term measured from the author's death, unless the concern is with the longevity of his heirs. The first point is mercantilist, reflecting the fact that the United States is a net exporter of intellectual property. One of the committee reports does explain, however, that the extension of the copyright term "will provide the important collateral benefit of creating incentives to preserve existing works [in digital format]." S. Rep. No. 315, 104th Cong., 2d Sess. 13 (1996). This is similar to our colorization example.

The mercantilist argument is not a good one even if a weight of zero is given to foreigners' welfare. Foreign trade is only a small part of the U.S. economy, and most intellectual property produced in the United States is consumed here. Moreover, U.S. producers may be hurt rather than helped by expanded copyright protection because it increases their input costs, as stressed in Chapters 2 and 3.

ex ante or ex post, though, as we have seen, ex post extension involves greater deadweight costs but lower rent-seeking costs. If the method of extension were periodic renewal, then since Congress could cut off the right of renewal at any time, a law authorizing renewals without a number limit would be less vulnerable to a constitutional challenge than a grant of perpetual copyright *ab initio*—the latter being flatly inconsistent with granting a copyright for "limited Times." Periodic renewal would also have the superior economizing properties that we have emphasized.

But it would not address the case in which intellectual property that has fallen into the public domain by abandonment is sought to be revived. Suppose Tobias Smollett had copyrighted his books but after a few renewals his heirs had decided the books had no value and so declined to pay the renewal fee. Our analysis implies that a publisher who wants to publish Smollett's books today should be permitted to take out copyright on them, by analogy to the rule discussed in Chapter 1 that allows finders to obtain title to abandoned (as distinct from merely misplaced) physical property. Unfortunately, the efficient implementation of such a rule would be considerably more complicated in the case of intellectual property. Allowing abandoned physical property to be withdrawn from the public domain unproblematically implements the policy that valuable property should *in general* be owned in order to create the correct incentives for its exploitation. But imagine a "finder" of intellectual property who claimed to have found, and who sought to obtain copyright in and register, all the books in the British Library on which copyright had expired. If the claim were allowed, the situation would resemble the banking of trademarks by listing all possible combinations of the letters of the alphabet, and would thus create the rent-seeking and transaction-cost problems that we discussed in Chapter 7 in relation to such banking.

The problem may not be insoluble. One way to deal with it would be to limit the acquisition of copyright in previously created works to works created after the law authorizing such acquisition was passed. Another would be to require the publication of such a work within a specified period of time after copyright was claimed in it. Another would be to charge a stiff fee for registering such a copyright.

We need to consider one more objection to indefinite renewals and to allowing even limited copyright protection for "found" public domain works. The expenditures that these measures would induce on discovering and disseminating obscure public domain works and on maintaining consumer interest in copyrighted works that were about to fall into the public domain would not be expenditures on creating expressive works; they would be marketing expenditures, which are not within the traditional domain of intellectual property law. Why should a particular subset of such expenditures be singled out for legal protection? Firms that introduce wine and cigar bars, recognize the potential for health clubs that combine workout and social activities, or introduce baggy trousers or pastel colors for clothing cannot prevent other firms from imitating their marketing innovations. Although trademark law may prevent confusingly close imitation and copyright law may protect particular advertising slogans, these laws do not prevent competitors from free riding on the information developed by the market innovators.

But the distinction between expressive works and the marketing of those works is overdrawn. Consider a record company that develops, promotes, and distributes new pop records. Which will be hits and which flops is not knowable in advance. In the absence of copyright protection—and here we are speaking just of protection against the copying of the sound recordings themselves—unauthorized copying could drive the price of the successful recordings down to their cost of manufacture and distribution and leave nothing for covering the costs of developing and promoting recordings of new songs and new performers. Copyright protection enables the record company to earn enough money on the hits to cover both the costs of the hits and the production *and marketing* costs of the many failures. By doing this, copyright indirectly prevents free riding on marketing expenditures similar to those incurred to maintain interest in soon-to-expire copyrighted works.[34] To state this another way, exploring the market for expressive works—the sort of exploration that the measures we are discussing would encourage—is a stage in the creation of intellectual property.

Moreover, marketing expenditures associated with expressive works differ from those associated with the other examples of new products that we gave because it is easier in the case of most expressive works to identify the innovator. Many people might be able to make credible claims of being the first to come up with the idea of baggy pants or pastel shades of clothing or combining exercise and social opportunities under the same roof. Having the legal system try to sort out these competing claims would involve substantial costs that would usually be avoided when someone was seeking to restore a copyright in an obscure public domain work, and the problem would never arise when one was seeking merely to renew an existing copyright.

Finally, many new business ideas may now be legally protectable by business-method patents (see Chapter 10). Such protection is often of marketing

34. This concern is explicit in the copyright statute, which discourages free riding on marketing and promotional expenditures of previously unpublished works about to fall into the public domain by conditioning an additional forty-five years of protection on publication. See 17 U.S.C. § 303(a). Thus copyright on works unpublished on January 1, 1978, continues until December 31, 2002, but if they are published by that date protection continues until December 31, 2047.

gimmicks, such as Amazon.com's "one-click shopping," and thus provides incentives for firms to invest in marketing and promotion that would be subject to free riding in the absence of legal protection.

Empirical Analysis

Data on copyright registrations and renewals over the past century are abundant[35] and enable us to add an empirical dimension to our analysis. If it turned out that all or most copyrighted works that antedate the 1976 Copyright Act were renewed, the implication would be that a system of indefinite renewals might approach perpetual copyright, though this would also depend on how steep the renewal fee was. Conversely, if renewals were infrequent even though renewal fees were nominal,[36] probably only a relatively few highly valuable works would remain under copyright beyond the initial term if indefinite renewals were permitted.

The number of registrations is only a proxy for the number of copyrighted works because registration is not a prerequisite for copyright protection. But both the 1909 and 1976 copyright acts created strong incentives to register a copyright and to register it promptly. Not only is registration (or, under the 1976 Act, an application to register) a prerequisite to filing a suit for infringement,[37] but it must be done before the infringement (or within three months of first publication) if the copyright holder wants to recover statutory damages and attorney fees. The 1909 Act fixed the copyright term at twenty-eight years from the date of first publication (or, for works that were copyrighted but not published, from the date of registration), and at the end of the term the copyright could be renewed for an additional twenty-eight years (raised to forty-seven years in 1962 and to sixty-seven years in October 1998) if the copyright holder applied for renewal within the last year of the initial term. Beginning in 1992 renewal became automatic,[38] so renewal registrations were sure to decline after that, but not to zero because there was still an incentive to file: a renewal registration is prima facie evidence of the validity of the copyright during its extended term and of the facts stated in the

35. See *Annual Report of the Register of Copyrights,* various years, published by the U.S. Copyright Office. See also Barbara A. Ringer, "Renewal of Copyright," in *Studies on Copyright: Arthur Fisher Memorial Edition,* vol. 1, p. 503 (1963 [1960]).

36. See note 8 above.

37. Under the Berne Convention Implementation Act of 1988, the registration requirement applies only to a "United States work"—a work first published in the United States or where the author is a U.S. national or lives here. See 17 U.S.C. § 101 for the complete definition.

38. Remember that these are renewals of copyrights that date from before the effective date of the 1976 Act, which gave the copyright owner a nonrenewable term of life plus fifty (later seventy) years.

certificate of renewal. Another statutory change, however, the extension of federal protection by the 1976 Act from published works to all works fixed in a tangible medium, could be expected to increase the number of registrations without any increase in the output of copyrightable works.

Our primary focus is on renewals because they allow us to estimate the expected economic life of a copyright. But we need data on registration as well because the number of initial registrations determines the number of works that are potentially renewable twenty-eight years later. For example, works renewed in 1938 were registered initially in 1910. To obtain the number of 1910 registrations, we have to deduct renewal registrations in 1910 (from works first copyrighted in 1882) because the Copyright Office includes renewal registrations in its tabulation of registration.

Figures 8.2 through 8.4 graph registrations, renewals, and the renewal/registration ratio (which in year t is simply the number of renewals in t divided by the number of initial registrations in $t-28$) over the past century.

Copyright registrations and renewals rose rapidly in the twentieth century, but, as expected, renewals began to decline in 1992 when they became automatic.[39] The rise doubtless reflects an increase in the number of copyrightable works brought about by a growth in the economy's total expressive output, as well as reflecting changes in copyright law. Why then did both registrations and renewals peak in 1991, declining by almost 20 percent by 2000, with the decline concentrated in the last year? The answer may be, in part anyway, that the registration fee doubled in 1991, from $10 to $20, and increased again in 2000, to $30, while the renewal fee doubled to $12 in 1991, rose to $20 in 1993, and more than doubled, to $45, in 2000. Although these fees seem small in relation to the inconvenience of registering and complying with other requirements of registration, such as submission of a copy of the work to the Copyright Office,[40] Figures 8.2 and 8.3 suggest substantial negative responses to higher fees for both original and renewal registrations. The

39. Recall that the 1909 Copyright Act (effective July 1, 1909) extended the renewal term from fourteen to twenty-eight years. Works that had been renewed for fourteen years in the period 1895 to 1909 were entitled to a further renewal of fourteen years, for a total renewal period of twenty-eight years. We do not include the fourteen-year extensions in our count of renewals in the 1910 to 1923 period. Our analysis of renewal data begins with the fiscal year ending June 30, 1910, the first fiscal year of the 1909 Act.

40. Another reason for the decline in registrations may be that since 1989 registration has no longer been a condition for bringing an infringement suit for foreign works protected by the Berne Convention and the World Trade Organization, though it remains a prerequisite for seeking statutory damages and attorneys' fees. The fact has little quantitative importance, however, because, as we'll see, foreign works are only a small fraction of copyright registrations. Also, statutory damages and attorneys' fees are significant remedies and so provide significant incentives to continue to register foreign works.

11

The Economics of Patent Law

The standard rationale of patent law is that it is an efficient method of enabling the benefits of research and development to be internalized, thus promoting innovation and technological progress. We suggest in this chapter that a more illuminating way of thinking about the patent system is as a response to economic problems inherent in trade secrecy and market structure.

The Economic Logic of Patents and Patent Law

PATENT VERSUS COPYRIGHT In light of the emphasis in so many of the preceding chapters on copyright law, we begin by noting the similarities between patents and copyrights—but also the important legal and economic differences. The conventional rationale for granting legal protection to inventions as to expressive works is the difficulty that a producer may encounter in trying to recover his fixed costs of research and development when the product or process that embodies a new invention is readily copiable. A new product, for example, may require the developer to incur heavy costs before any commercial application can be implemented, so that a competitor able to copy the product without incurring those costs will have a cost advantage that may lead to a fall in the market price to a point at which the developer cannot recover his fixed costs.

A twist not present in the copyright area (with some exceptions noted later) is that in the absence of legal protection for an invention, the inventor will try to keep the invention secret, thus reducing the stock of knowledge available to society as a whole. Patent law combats this incentive by requiring, as a condition of the grant of a patent, that the patent application (which becomes a public document if and when the patent is issued, and often in any event, as we'll note in Chapter 13, eighteen months after filing) disclose the steps constituting the invention in sufficient detail to enable readers of the application, if knowledgeable about the relevant technology, to manufacture

the patented product themselves. Of course they may not use the information to make or sell or use the patented product or process itself without a license from the patentee. But they may use it for any other purpose, including attempting to "invent around" the patented invention—that is, to achieve the technological benefits of the patent without duplicating the particular steps constituting it and thus without infringement. Inventing around facilitated by the required disclosure of the patented invention is a limitation on the monopoly power that the patent right confers. Patent law often yields broader protection than trade secrecy in respects that will become clearer in Chapter 13 and saves the inventor the cost of keeping his invention secret. But in return it exacts disclosures that facilitate inventing around. The swap is not advantageous for all inventors, which is one reason why trade secrecy abounds even in domains of inventive activity in which patent protection is obtainable. Another reason is that by teaching how to make the invention, the required disclosure teaches would-be infringers how to infringe, creating a risk that the patentee will have to defend the validity of the patent in expensive litigation.

A patent is more difficult to obtain than a copyright because a lengthy (though, as we'll point out in the next chapter, not very demanding) examination of the patent application by the Patent and Trademark Office is required, and because the drafting of patent claims that will at once withstand a court challenge to their validity yet not be so narrowly drawn as to offer little protection against competitors requires considerable legal skill. But at the same time a patent can confer greater value on its owner than a copyright does, even though the term of a patent is much shorter (its greater breadth is a reason *why* the term is shorter) and even though patent protection does not extend to the patent counterpart of the derivative work, which is the improvement patent. One reason patent protection can be more valuable than copyright protection is that a patent protects against any duplication of the patented invention rather than merely forbidding the copying of it. As we noted in Chapter 4, it usually is feasible to search the patent register for previous inventions, but it is not feasible to search the Library of Congress for previously copyrighted works that a new work might duplicate, especially now that a copyrighted work need be deposited in the Library of Congress only if the owner wants to register his copyright. But there is another important difference. Simultaneous or nearly simultaneous discovery or invention is much more common in the case of ideas, which are (more precisely, a subset of which are) what patents protect from duplication, than in the case of expression, the domain of copyright law. That is why there are patent races but not copyright races. If patents did not protect against independent duplication,

an inventor who had spent enormous sums to be the first to discover some useful new idea might find himself unable to recoup his costs because someone else, working independently toward the same goal, had duplicated his discovery within weeks or months after he made it. Patent law prevents such disappointments but at the same time, of course, fosters patent races and the rent-seeking costs that such races can impose.

Another reason patent protection tends to confer greater value on the patentee than copyright protection confers on the copyright holder is that, as just mentioned, patents protect ideas, which have potentially vast commercial application, rather than protecting merely a particular verbal or aural or visual configuration of ideas that are in the public domain—for remember that copyright law does not forbid copying the ideas in a copyrighted work.

With greater legal protection for patentees than for copyright holders comes a greater danger that the inventor will be enabled to charge a higher price than he needs to recover the fixed costs of his invention, thereby restricting access to the invention more than is necessary. This fear may explain why the term of a patent has always been shorter—it is now much shorter—than the term of a copyright, although proper comparison requires discounting to present value. At a discount rate of 10 percent, the present value of a constant stream of income to be received for twenty years[1] is 85 percent of the same stream received in perpetuity, so that in effect the patent statute allows the patentee to realize a maximum of 85 percent of the market value of his invention. At the same discount rate the present value of a constant stream of income to be received for 110 years (the copyright term if the author was forty years old when he created the copyrighted work and died at eighty) is 99.997 percent of what it would be if the term were unlimited.

These calculations are sensitive to the choice of discount rate[2] and to the assumption that the stream is constant forever. But if, as would be more realistic, we assumed that the stream of income would decline after a time—in other words, assumed that the patent or copyright would depreciate, as we assumed for copyrights and trademarks in Chapter 8—the percentages just calculated would be even higher. Later in this chapter we estimate the depreciation rate of the average patent to be 6 percent. If this figure were plugged into the present-value formula, the twenty-year patent term would yield almost 95 percent of the value of the patent in perpetuity at a 10 percent discount rate. Granted, the 95 percent estimate is too high because the twenty years run from the date of application, not the date of issuance, of the patent. Some patents will get about nineteen years of protection, but others only about fifteen; the average is likely to be eighteen (see note 1). If we substitute eighteen for twenty years, the 95 percent estimate falls—but only to 93 percent.

1. Since 1995 the patent term has been twenty years from the date of the patent application. Before then, and going right back to the first federal patent statute, enacted in 1790, the term had been seventeen years from the date the patent was granted. Since patent applications pend in the Patent and Trademark Office for anywhere from one to five years (sometimes, though rarely, longer), it is not certain that the change in the length of the term actually enlarged patentees' rights. However, a study of patents issued under the old law concludes that the patentees would have received on average a net increase of 253 days of patent protection had the new law been in effect instead. See Mark A. Lemley, "An Empirical Study of the Twenty-Year Patent Term," 22 *American Intellectual Property Law Association Quarterly Journal* 369, 385 (1994). See also id. at 392.

2. At a discount rate of 5 percent, the percentage figures in the text fall to 62.3 and 99.5. But the higher discount rate seems more appropriate in light of the uncertainty associated with income from intellectual property.

A FORMAL ECONOMIC MODEL How far the law should go in protecting an inventor from competition and the related question of how patents affect the incentives to invent can be illuminated with the aid of a simple model. We illustrate with a process patent that reduces the cost of making an existing product (X) rather than with a patent on a new product. The model could easily be extended to a patent on a new product, however, since the new product can be analyzed as the special case of an invention that brings the cost of a new product down from a prohibitive level (that is, at which the supply curve lies everywhere above the demand curve) to a level at which the product can be profitably produced.

In Figure 11.1, DD is the demand curve for X, MR the marginal revenue curve, MC_1 the industry marginal cost or supply curve before the cost-reducing innovation, and MC_0 the marginal cost curve of the firm (assumed to be constant throughout the range of feasible outputs) that develops the cost-reducing invention. The difference between the two marginal cost curves is the cost savings at each unit of output brought about by using the new technology. To simplify, we assume that many firms can produce X at MC_1 (implying a perfectly elastic industry supply curve equal to MC_1), yielding the pre-invention equilibrium outcome of price P_1 and output X_0. The conventional analysis of patents implies (with a limited exception noted below) no significant change in industry output or price as a result of the process patent. (In contrast, a product innovation will always lead to a greater output, since the pre-innovation output is zero.) Instead the patent holder maximizes his profits per period[3] (the cross-hatched area in the diagram) either by produc-

3. The diagram depicts only one period. The full profit from (also social gain from, as noted below in the text) the invention is the present value of the cost savings in the current and future periods minus the cost of the invention.

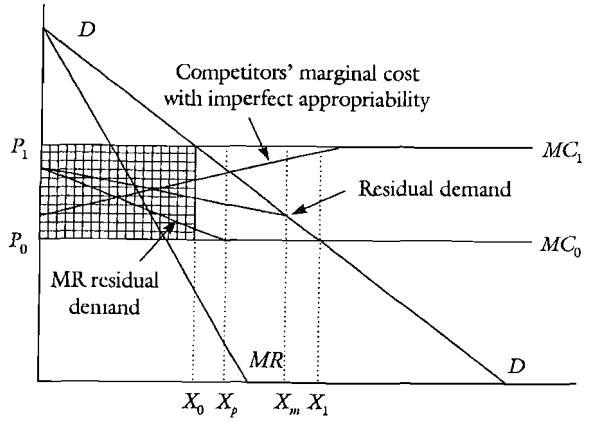

Figure 11.1. Effect of patent protection on copying

ing the entire industry output X_0 at a price slightly below P_1 or by licensing the patent to existing firms at a royalty rate just below $P_1 - P_0$.[4] Although output and price are essentially unchanged, the invention still yields a net social gain equal to the cost saving (the cross-hatched area) minus the cost of developing the invention. If, however, the demand elasticity at the competitive output is sufficiently large, marginal revenue will be above P_0 at X_0, which will lead the patent holder to cut price below P_1 and expand output beyond X_0. In general, the more elastic demand is at X_0 and the bigger the cost reduction resulting from the invention, the more likely price is to be below P_1 and output above X_0.

The conventional analysis that we have just sketched overlooks a key feature of patent law, however, by assuming that the patented innovation has no impact on the cost structure of other firms in the industry. As we noted earlier, the law requires that the patent application disclose the innovation in sufficient detail to enable persons of ordinary skill in the relevant technology to practice the invention. (Not that they are allowed to, of course, while the patent is in force, without the patentee's permission.) Once the patent is issued, and often earlier, the application is publicly available, and the information contained in it will help competing firms to invent around the patent and thus lower their costs of production without infringing the patent. (Presumably, the more information a firm uses from the patent application, the lower its costs but the greater the risk of being sued for infringement.) In addition to facilitating inventing around, the information in the patent application will help the competitor to determine the feasibility of a process improvement without having to incur the cost of experimenting to obtain this information. It may even enable him to get away with a certain amount of infringement. The patent holder may be reluctant to sue for infringement because of the risk that the court will hold that the patent is invalid. He may therefore tolerate modest levels of infringement.

4. The patent holder equates marginal cost to marginal revenue, which coincides with P_1 up to X_0 and with MR at outputs greater than X_0. At X_0 marginal revenue is discontinuous, because it equals P_0 just before X_0 and MR just after X_0.

The requirement of public disclosure creates a situation of *incomplete appropriability* by the patent holder, which in Figure 11.1 will cause the marginal cost curve of the patentee's competitors to fall below MC_1. Of course it is unlikely to fall all the way to MC_0, the patentee's marginal cost. In addition, it will probably become upward sloping rather than perfectly elastic as it falls because firms will differ among themselves in the costs of inventing around the patent, infringing the patent without being sued, and so on (recall the parallel example of copiers' upward-sloping supply curve in Chapter 3). If the competitors are price takers and the patentee acts as a dominant firm, the patentee will face a downward-sloping residual demand curve equal to the difference between DD—the overall demand for X—and the quantity that would be supplied by competitors at every different price of X. The patentee maximizes profits by equating MC_0 to the marginal revenue derived from his residual demand curve, yielding in equilibrium a total output of X_m consisting of the patentee's output of X_p and the competitors' output of $X_m - X_p$. The market price will lie between P_1 (the pre-invention price) and P_0 (the patentee's marginal cost). In short, output will be greater than the pre-patent output of X_0 and price will be lower than P_1. This contrasts with the conventional analysis, in which output and price before and after the innovation are (with the qualifications noted) unchanged.

The greater patent protection is, the smaller the benefit to competitors from the information contained in the patent grant because the less they can do with it. They will face greater difficulty and higher costs in inventing around the patent, a higher probability of losing a patent infringement suit, and greater sanctions if they lose. Their marginal cost curve will be steeper or, equivalently, the elasticity of supply lower, making the residual demand curve less elastic and thus enabling the patentee to charge a higher price and capture a greater share of the post-invention market.[5]

5. The elasticity of the residual demand curve can be written as $\varepsilon_r = \varepsilon_d/s + \varepsilon_s(1 - s)/s$, where ε_d is the elasticity of the demand curve, ε_s the elasticity of the competitors' supply curve, and s and $1 - s$ the patentee's and the competitors' output shares, respectively. The greater are ε_d and ε_s and the smaller is s, the greater ε_p will be. The patentee's profit-maximizing ratio of P/MC_0 is $[\varepsilon_d/s + \varepsilon_s(1 - s)/s]/[\varepsilon_d/s + \varepsilon_s(1 - s)/s - s]$ and thus is greater the lower ε_s and ε_d are.

Whether a given degree of patent protection is socially desirable depends on the patentee's fixed costs, the inherent difficulty of inventing around the patent (that is, holding constant the degree of patent protection), and the extra profits that the patentee can expect to receive from greater protection. The greater the fixed costs of research and development and the easier it is to invent around the patent, the greater will be the degree of patent protection required to create adequate incentives to invest in developing the invention in the first place. The patent system makes no effort, however, to match the degree of patent protection to those variables. A patentee's monopoly markup, which of course is influenced by the degree of patent protection, bears no direct relation to the fixed costs that he actually incurred in creating the patented invention. If the elasticity of demand is low because competitors find it very difficult to invent around or, in the case of a new product, because the product provides benefits to consumers not available from other products, the patentee will be able to charge a high price relative to marginal cost and thus obtain revenues that may greatly exceed what may be modest total costs. Not only will access to the patented invention be restricted to a greater degree than would be necessary to create incentives for optimal invention, but the prospect of such windfalls will induce rent-seeking behavior, with a resulting waste of resources illustrated by patent races, which are discussed below. Copyright has no direct analog to a patent race, but there is an indirect one: the "dry hole" phenomenon, discussed in Chapter 2, that may lead to the overproduction of books and other expressive works.

The length of a patent illustrates the potential disjunction between actual and optimal patent protection. If as we suggested earlier a patent enables the patentee to retain 85 percent of the profit generated by it even if there is no depreciation, the twenty-year patent term is excessive if a perpetual patent would be expected to overcompensate the patentee by more than 17.6 percent (.15 ÷ .85).[6] If we factor in depreciation at 6 percent, the twenty-year term is excessive if a perpetual patent would overcompensate the patentee by more than 5.3 percent (.05 ÷ .95). There is no theoretical or empirical basis for supposing that any of these figures approximates the level at which the average patentee would recover the cost of invention but no more.

PATENT RACES A patent race is a race among competing firms to be the first to discover and patent some new idea having commercial potential. Such

6. See George J. Stigler, "A Note on Patents," in Stigler, *The Organization of Industry* 123 (1968). (For example, if the market value of the patent if perpetual is $1 million and the patentee should have received only $850,000, $150,000 represents excess compensation and it is 17.6 percent of the $850,000 that he should have received.) The percentage in the text rises to 25 percent under the seventeen-year term in force before 1995 if the difference between the date of patent application and the date of grant is disregarded.

a race can generate costs of invention that exceed the social benefits,[7] because the first competitor to reach the finish line will obtain the patent and thus the full value of the invention (minus the part that is externalized by the limitation on the length of the patent term, but we have seen that under plausible assumptions that part is small) even if he beat his competitors by only a day. Suppose the present discounted value of the patent is $10 million, three firms are in the race, and if each spends $2 million each will have a 33 percent chance of being the first to invent. Assume further that the social surplus is maximized if each firm spends just $2 million on trying to come up with the invention. From a private standpoint, any of the competitors that thinks that by spending another $5 million on research and development he can be the first to come up with the invention by even just one day may have an incentive to make that expenditure. We say "may" rather than "will" because the decision will depend on the competitor's expectations concerning the likely reaction of the other firms. He may fear that they will counter his expenditure of $5 million with expenditures of their own that will cause his expenditure to be a poor investment. On the other hand, he may think he can conceal the expenditure from his rivals. Among still other possibilities, each of the competitors may invest a smaller amount in the race or one may drop out of the race and each of the others spend $5 million. In any event, it is quite possible that more than the optimum total of $6 million (3 × $2 million) will be spent on R&D. This assumes that the additional costs, unless trivial, incurred in trying to beat rivals by one day to the patent finish line will exceed the social benefits of having come up with the invention one day earlier, but this is a plausible assumption—especially since the earliest inventor is not necessarily the one who can produce the product or process embodying the invention the fastest. He may be better at inventing than at producing. If so, after winning the race he may decide to license the fastest or most efficient producer, in which event the transaction cost incurred in the licensing process will be still another cost of the race.

Two qualifications to the economic criticism of patent races should be noted. First, the research expenditures by the losers of the race may not be wasted even if the race does not accelerate the inventive process by a day, for the expenditures will generate information that the losers may be able to use in other projects. Second, patent races need not produce any social waste at all in cases, which are particularly common in the pharmaceutical drug industry, in which there are as it were multiple prizes and hence more than one winner. For example, SSRIs (selective serotonin reuptake inhibitors), such as

7. See, for example, Dennis W. Carlton and Jeffrey M. Perloff, *Modern Industrial Organization* 522–526, 536–540 (3d ed. 2000), summarizing an extensive literature illustrated by Partha Dasgupta and Joseph Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D," 11 *Bell Journal of Economics* 1 (1980).

Prozac, Zoloft, and Paxil, are competing antidepressant drugs, but they are based on different chemicals and so are separately patentable.

RULES THAT REDUCE THE SOCIAL COSTS OF PATENT PROTECTION Patent law employs a number of devices to minimize social costs, besides the previously discussed requirement of public disclosure. The patent's limited term is, of course, one of the devices. This limitation may, as we have pointed out, be largely illusory because of discounting to present value and its interaction with depreciation. But not completely. A byproduct of patent racing is that patents frequently are obtained well before commercial development is complete (famous examples are the jet engine, radio and television, and fluorescent lighting), which reduces the economically significant length of the patent term below twenty years. Unlike a copyright or a trademark, a patent can be obtained before there is a working model of the patented product or process—it's as if one could obtain a copyright on a book on the basis of an outline, or a trademark without having a product to sell. The early grant of a patent not only shortens the time in which the inventor can charge a monopoly price because the patent term runs from the application for or (before 1995) from the grant of a patent rather than from the commencement of production and sale; it may also minimize wasteful duplication of effort by competing inventors.[8] The early grant serves the same purpose as privatizing a common pasture or as the committed-searcher doctrine, which allows the first searcher who commits himself to search for an abandoned piece of property to prevent others from obtaining title to it for as long as his search is being conscientiously pursued (see Chapter 1). But the doubt expressed when we said "*may* also minimize wasteful duplication" bears emphasizing; like so many other propositions about patent law, this one can be questioned—and will be shortly.

Inventions are not patentable unless they are useful, novel, and nonobvious. Let us try to give economic meanings to these terms. The requirement of utility can be understood to have three economic purposes. One is to rule out patents on basic research, and another is to delay the point in the development of a new product or process at which a patent may be obtained; we'll return to both these points. The third is to reduce the cost of patent searches by screening out useless inventions by cranks or amateurs, or by inventors hoping to blanket an area of research with patents in the hope of forcing researchers who come up with useful inventions within the area to seek licenses from them. In other words, the requirement of utility serves to

8. This is the "prospect theory" of patent law, proposed in Edmund W. Kitch, "The Nature and Function of the Patent System," 20 *Journal of Law and Economics* 265 (1977).

limit strategic patenting—a serious problem of the patent system, as we shall note. An alternative to requiring proof of utility, however, would be to increase the patent fee.

The requirement of novelty prevents getting a patent on something known to have been invented already. Unlike many foreign patent systems, the U.S. system gives the right to patent to the first inventor (with a qualification noted later) rather than to the first filer of a patent application. The effect is to economize on search costs, broadly defined. There are two ways of searching for an idea that one might patent. One is to search the prior art, including unpatented as well as patented inventions, to see whether the invention has already been made. That requires a search of technical articles as well as of the patent registry. The other method is to search just the patent registry and if one doesn't find a patent on the invention, to conduct the necessary R&D. The second method is more costly in those cases in which an invention is part of the prior art but not already patented; for there is wasteful duplication in that case. The U.S. rule encourages the first method of search, since if the invention is in the prior art even though it is not patented, an application for a patent will fail for want of novelty.[9]

There is an ambiguity in the concept of novelty that is related to the discussion in Chapter 1 of abandonment. We are not interested in the metaphysical question whether ideas, lacking as they do spatial or temporal bounds, can be said to have existed before they were discovered and therefore to lack novelty: it has always been the case that a certain sequence of steps would enable a heavier-than-air structure to fly. Our concern is with the practical question whether property rights should be recognized in otherwise patentable inventions that have been abandoned. The patent statute is not much help here. The requirement of novelty bars the grant of a patent if before the date of the invention sought to be patented the invention was "known or used by others."[10] Read literally this would bar the grant of a patent on an abandoned invention. But the Supreme Court long ago held that if knowledge of an invention has been completely lost, the re-inventor can patent it.[11] In addition,

9. See Robert Patrick Merges and John Fitzgerald Duffy, *Patent Law and Policy: Cases and Materials* 419–422 (3d ed. 2002).

10. 35 U.S.C. § 102(a). We examine in Chapter 13 the significance of this requirement for the patentability of an invention that someone else made and used earlier but kept as a trade secret. Use is vital. Under the U.S. first-to-invent rule an inventor does not forfeit his right to patent his invention by virtue of not having filed a patent application promptly after having made the invention, though if he publishes the invention he must then apply for a patent on it within one year.

11. Gayler v. Wilder, 51 U.S. (10 How.) 477 (1850). See also Corona Cord Tire Co. v. Dovan Chemical Corp., 276 U.S. 358 (1928); Allen v. W. H. Brady Co., 508 F.2d 64 (7th Cir. 1974).

abandonment can be evidence that the first inventor was unable to reduce his idea to practice and so was not really the first inventor in the sense relevant to patent law; he failed the test of utility.

Were "nonobviousness" interpreted literally, it would add little to the requirements of utility and novelty, since if an invention is both useful and obvious, why hasn't it been discovered already? Maybe because an unexpected shift in demand and supply has suddenly made it useful and *someone* has to be the first to grasp the fact; but that is a special case. A more illuminating approach ties nonobviousness to uncertainty and cost.[12] Invention is a matter of adding to the stock of useful knowledge and so of reducing uncertainty. What is already known is not something waiting to be invented. But sometimes an idea is unknown not because it would be costly to discover but because it has no value. If some exogenous shock gives it value, it will be discovered more or less simultaneously by a number of those who can exploit it; there is no need to give exclusive rights to the first discoverer. But if it is costly to dispel uncertainty, then since the cost is incurred before a product embodying the invention can be brought to market, competitors will be tempted unless blocked by patents to sit back and wait until the invention is made and then sell copies, thus free riding on the inventor's cost of invention. Uncertainty and cost interact, in other words, as we also noted in regard to expressive works in Chapter 2. Uncertainty implies the likelihood of failures en route to success. Those failures are costly, and since the costs are incurred before the successful invention can be patented and marketed, they are additional fixed costs that the inventor must recover in the revenues generated by his patent.

Uncertainty has a further significance. In his classic article on the economics of invention, Kenneth Arrow pointed out that risk aversion would result in underinvestment, from a social standpoint, in risky undertakings, such as invention.[13] This point balances Arnold Plant's argument that patentability draws resources from what might be socially more valuable productive activities that do not offer monopoly returns. Unfortunately, the weights of these two offsetting factors are unknown.

To the extent that the requirement of nonobviousness succeeds in preventing the patenting of inventions that would not cost a lot to discover and perfect, it limits patent races, which are more costly the greater the net expected gain from winning. There is also an evidentiary argument for the requirement of nonobviousness. Inventions that are obvious are likely to mark only small advances over the prior art. This will make it difficult, when a patent expires, to determine whether someone who appears to be using the invention covered by the expired patent may actually be infringing a later patent, not yet expired, that made a tiny improvement over the invention covered by the expired patent. Here the requirement of establishing that one's invention is not obvious serves the same evidentiary function that we identified in Chapter 4 for the requirement that a derivative work of a copyrighted work, to be copyrightable itself, have significant originality.

There is a growing tendency, fostered by the Federal Circuit (the court with exclusive jurisdiction over patent appeals—see Chapter 12), to use commercial success as a proxy for nonobviousness. The theory is that if an invention is both obvious *and* lucrative, why wasn't it thought of earlier? The tendency has been criticized for failing to distinguish between invention itself and marketing, the latter involving inputs other than the invention that are not protected by intellectual property law.[14] Courts cannot readily disentangle the contribution of the invention to the commercial success that attends its marketing. It is odd, though, that use of commercial success as a proxy for nonobviousness should be encouraged by the Federal Circuit, since evidence of commercial success is the sort of evidence that courts with no technical knowledge feel comfortable with, and the Federal Circuit has that knowledge, or at least to a greater extent than other U.S. courts do. It is a further indication that, as we shall see in the next chapter, the Federal Circuit has a bias in favor of patentability.

An important limitation of patent law is that fundamental ideas, such as physical laws, cannot be patented. The domain of this rule is different from that of the parallel limitation in copyright law. The "ideas" found in an expressive work that are ineligible for copyright protection are standard plots, stock characters, verse forms, literary and musical genres, schools of painting, dramatic conventions, iconography, and the like. The ideas that patent law excludes are fundamental scientific (including mathematical) and technological principles. The two classes of idea (call them "expressive" and "inventive") are related, however, both in the enormous potential for rent seeking that would be created if property rights could be obtained in them and in the enormous transaction costs that would be imposed on would-be

12. On the cost element, see Edmund W. Kitch, "*Graham v. John Deere Co.*: New Standards for Patents," 1966 *Supreme Court Review* 293; Roberts v. Sears, Roebuck & Co., 723 F.2d 1324, 1344 (7th Cir. 1983) (en banc) (separate opinion). On the uncertainty element, see Robert P. Merges, "Uncertainty and the Standard of Patentability," 7 *High Technology Law Journal* 1 (1992).

13. See Kenneth J. Arrow, "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors* 609, 610–614 (1962).

14. See Robert P. Merges, "Commercial Success and Patent Standards: Economic Perspectives on Innovation," 76 *California Law Review* 805 (1988).

users.[15] The transaction costs would be enormous because the scope of either kind of idea, the expressive or the inventive, often is extremely difficult to pin down, and this would make it difficult for newcomers to know when they needed to get a license. Apart from uncertainty, the more elements out of which a new expressive work or new invention could be made that were owned by someone (in other words, the more the public domain shrank) and therefore had to be licensed by the newcomer, the greater would be the transaction costs that he would incur.

The shorter the patent term, the lower the social costs of allowing fundamental principles to be patented. Basic research is distinguished from applied research mainly by lacking *immediate* commercial applications. Hence if the patent term were very short, the social costs of allowing the patenting of basic research findings would be minimal—though so would be the incentive for such research that patentability would impart. The patent term is not short, and the interval between basic research and commercial applications is shrinking;[16] so making basic research findings patentable could impose heavy social costs. The shorter the interval, however, the harder it is to deny patents on basic research. If commercial applications of some piece of basic research are immediately foreseeable, a patent on the research will pass the test of utility.

The distinction between fundamental idea and patentable invention is being further eroded by the rise of the business-method patent, for which such fundamental-seeming algorithms as the Black-Scholes options-pricing model would have been a plausible candidate had it been invented after the new type of patent was recognized.

An enormous amount of basic research is produced every year in the United States and other advanced countries without benefit of patentability. This is not in itself a compelling reason against extending patent protection to basic research, because it overlooks the role of government in funding such research. In 1999 half of all basic research in the United States was funded by the federal government, and of the balance 29 percent was financed by universities and other nonprofit research establishments out of their own funds.[17] In effect, basic research is incentivized by a reward system

15. There is also a considerable area of overlap between copyright-related ideas and patent-related ideas that is due to the many important technological innovations in the expressive media, including new methods of painting, color photography, and special effects in movies.

16. See Robert P. Merges, "Property Rights Theory and the Commons: The Case of Scientific Research," *Social Philosophy and Policy*, Summer 1996, p. 145 (1996).

17. Computed from NSF/SRS, *National Patterns of R&D Resources: 2000 Data Update*, tab. 2: National Expenditures for Basic Research from Funding Sectors to Performing Sectors—1993–2000, *http://www.nsf.gov/sbe/srs/nsf01309/start.htm*.

that involves prestigious academic appointments, lecture fees, grants that reduce teaching loads, and the prospect of Nobel and other prizes, while applied research (including, however, instruments and other tools used to conduct basic research) is incentivized by intellectual property rights. If patent protection extended to basic research, government would reduce its funding, taxes would be lower, and the allocative distortions caused by taxes would be smaller. However, a great deal of this research has no near-term commercial application and so could not be financed by patenting. In addition—and in tension with the point made above that patents on basic research would generate disproportionate rewards—because basic research has commercial value only as an input into further (applied) research activity, it would be very difficult to calculate license fees for patents on basic research, in which event patents might not be an effective method of eliciting such research.

Moreover, universities have strong incentives to support basic research—the leading universities are called "research universities" and regard the conduct of basic research as their main mission—while a system of government grants for basic research that supplement university resources operates better than it would in the commercial arena because politicization of basic research is less likely. Money for strictly scientific research conducted in universities is more likely to be allocated in accordance with objective scientific criteria. A scheme of financial support for commercial innovations would undoubtedly attract strong pressure from interest groups eager to persuade the government to give this or that industry a leg up by financing its R&D. Basic research by definition does not have immediate commercial applications, so it is less attractive to business groups and therefore they can be expected to exert less pressure on the government agencies that fund such research. Less is not zero; and universities themselves engage in lobbying, a form of rent seeking, for government grants.

Reference to basic research brings out a contrast between the role of nonpecuniary incentives in the creation of expressive works and in the creation of inventions. The principal rewards of aesthetic achievement flow to the authors (composers, painters, etc.) of the expressive works themselves rather than to the creators of the "ideas" reflected in them. Odd as it may seem, not much celebrity attaches to being the inventor or discoverer of perspective, the fugue, the sonnet, the obtuse narrator, the rhymed couplet, the opera, and so forth, as opposed to the perfecters of the form. How many people know that Monteverdi invented the opera? Even among those who do know this, he is far less celebrated than such operatic composers as Mozart, Wagner, and Verdi. The situation is the opposite in scientific and technological fields. There fame, a potent motivator with often a cash value to boot,

goes to the discoverer of basic ideas rather than to the individuals who perfect their application.[18] This is an argument for providing greater legal protection and therefore economic rewards to applicators than to discoverers in the scientific and technological as opposed to the cultural domain, and that is approximately the line drawn by the patent law. However, the line is eroding. As we mentioned, the shortening of the time between scientific discovery and technological application is, among other factors, enabling more and more fruits of basic research to be patented.

The nonpatentability of basic ideas is related to the distinction that patent law draws between discovery of that which has always existed and invention, denying patent protection to the former.[19] Superficially the relation is that basic ideas fit the concept of discovery better than that of invention; one supposes that $E = mc^2$ is a discovery about the structure of matter rather than an invention of Einstein's, although some philosophers of science would disagree. But the real point, which takes us back to the continental-discovery example of Chapter 1, is that when something is known to exist and is just waiting to be found, the danger of a wasteful race to find it is increased because the probability of success, and hence the expected gain, is greater.

The limitations that patent law imposes on the obtaining of a patent are procedural as well as substantive. In particular, one can't just assert a patent, as one can a copyright or trademark. One must submit a patent application to the Patent and Trademark Office (PTO) for determination whether the invention satisfies all the preconditions for patentability. The alternative would be to allow an inventor to assert a patent and let the courts decide whether the patent was valid. In fact the courts *do* decide, since the grant of a patent by the PTO creates only a rebuttable presumption of the patent's validity. So the difference between the patent and copyright regimes is that the former has two tiers of review and the latter just one. Even if the patent statute were amended to eliminate the preliminary review function of the PTO, there would doubtless still be a patent registry and patent searches so that inventors wouldn't waste their time on inventions that would not stand up in court if challenged. That is why the PTO also maintains a registry of trademarks, even though applications for trademark registration do not receive in-depth review.

The argument for administrative scrutiny of patent applications is that without it the patent registry would be clogged by patents of very dubious

18. On the motivations of scientists, see Paula E. Stephan, "The Economics of Science," 34 *Journal of Economic Literature* 1199, 1201–1203 (1996).

19. See Robert P. Merges et al., *Intellectual Property in the New Technological Age* 146–149 (2d ed. 2000).

validity asserted often for anticompetitive purposes. It *was* clogged between 1793 and 1836 when there was no patent examination—an experiment that failed in part because of patent registrations seemingly designed to extract rents from serious inventors.[20] This is not a *general* problem with copyright, though the qualification is important. The narrow scope of a copyright makes it infeasible for a publisher, say, to throw a monkey wrench into the business of a competitor by registering copyright on a host of books that have no commercial value but might occupy a competitor's entire field of prospective operation so that the competitor's efforts to exploit the field would be blocked by copyright; it also frustrates collusion between competitors to cross-license copyrights. There are exceptions: only a limited number of pleasing tunes can be constructed from a short series of notes,[21] and only a limited number of short sequences of computer code is possible. In both these cases it might be feasible to obtain, and at rather little cost, an immense number of copyrights that would tend to block subsequent composers and software writers. The law has been alert to the problem in the computer software context.[22] Recall from Chapter 4 the holding of the *Sega* case that the fair use privilege allows software to be copied without the copyright holder's permission when the copying is necessary for reverse engineering the software to generate the same functionality (unprotected by copyright law) by means of a different sequence of code. In the case of musical copyrights, the plaintiff must show that the defendant had access to the plaintiff's work and that the two works are substantially similar. And even if that is the case, the defendant can rebut the inference of copying by showing that he independently created the work. That showing is easier to make the more limited the alternatives open to composers, so that, unless a musical work that the com-

20. The Senate report accompanying the bill to reinstate patent examinations contains a vivid account of the country being "flooded with patent monopolies, embarrassing to bona fide patentees, whose rights are thus invaded on all sides . . . Out of this interference and collision of patents and privileges, a great number of lawsuits arise . . . It is not uncommon for persons to copy patented machines in the model-room; and, having made some slight immaterial alterations, they apply in the next room for patents." S. Rep. No. 338, 24th Cong., 1st Sess. 3 (1836).

21. In a sequence limited to four notes in a major scale that are then repeated four times, the total number of combinations is more than 4,000, but not all of these combinations will yield tunes sufficiently pleasing to have commercial value.

22. The "cover" and "jukebox" rules for musical copyrights, see 17 U.S.C. §§ 115–116, are addressed to a different problem. The cover rule allows performers to record copyrighted works without the copyright owner's permission, once he has authorized a distribution of his work in recorded form. The jukebox rule allows jukebox owners to place whatever recordings they want in their jukeboxes without permission of the copyright owners. These are examples of compulsory licensing, and in both cases the copyright owner is entitled to a fee.

poser is alleged to have infringed is widely performed, he is likely to have a strong defense of independent creation.

The administrative review process in the Patent and Trademark Office is lax (see Chapter 12). The consequence may be the creation of "patent thickets" that retard innovation, as we shall point out shortly.

Is Patent Law Socially Cost-Justified?

The most important economic question about the patent system is whether on balance, with the various twists and turns that we have mentioned, it increases or reduces economic welfare. Although there are powerful economic reasons in favor of creating property rights in inventions, there are also considerable social costs and whether the benefits exceed the costs is impossible to answer with confidence on the basis of present knowledge. The *relative* quality of patents can be estimated because patent applications are required to cite prior patents of which the applicants are aware on which the current application is based, and so the number and character of the citations to a given patent, much like the number and character of citations to a judicial opinion or a scholarly article, can be used as a proxy of quality.[23] Citation data can also be used to evaluate some policy issues relating to the inventive process. For example, the federal government has for a number of years now been encouraging its research laboratories to focus more on research having commercial applications.[24] One would like to know whether the new policy has been effective—and a study has found that government research is indeed

23. See Bronwyn H. Hall, Adam B. Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools" (National Bureau of Economic Research Working Paper No. 8498, Oct. 2001).

24. This is part of a broader program of encouraging the commercial exploitation of government-conducted or -supported research. Another part of the program, the Bayh-Dole Act of 1980, authorizes universities and other research entities to patent the fruits of research supported by federal funds. See Rebecca S. Eisenberg, "Public Research and Private Development: Patents and Technology Transfer in Government-Sponsored Research," 82 *Virgina Law Review* 1663 (1996). Previously such fruits were in the public domain. A foreseeable and, it appears, an actual effect of the change in policy has been to provide windfalls to universities in the form of royalties for inventions financed in large part by the federal taxpayer and to encourage universities to shift their research emphasis from basic to applied research. See id. at 1708–1714; Peter S. Arno and Michael H. Davis, "Why Don't We Enforce Existing Drug Price Controls? The Unrecognized and Unenforced Reasonable Pricing Requirements Imposed upon Patents Deriving in Whole or in Part from Federally Funded Research," 75 *Tulane Law Review* 631, 668 (2001), and references cited there. That shift in emphasis shows, however, that patentability does affect research incentives. The effect is complex. By increasing the university's revenues, the patenting of the fruits of applied research generates funds that can be used to support basic research, and by providing additional income to university researchers may avoid losing them to industry.

being cited more frequently in private patents, which suggests that the answer is yes.[25]

But citations to patents do not reveal whether an invention would have been made without the prospect of its being patented. Even assessments of the relative value of patents on the basis of citations are of limited reliability, especially if social rather than private value is the concern. The usual method of assessment is to deem a patent more valuable the more various the areas of technology it is cited in. But the broader a patent's reach, the more likely it is both to impede inventive activity by increasing the likelihood that a new invention will infringe an existing patent (the same concern we expressed in Chapter 2 with regard to broad copyright protection) and to cover basic research.

Patent "renewal" rates are another potentially illuminating but ultimately inconclusive body of data.[26] We use scare quotes because a U.S. patent cannot be renewed after the expiration of its twenty-year statutory term. To keep it in force for the full twenty years, however, the patentee must pay maintenance fees of $880 at 3.5 years, $2,020 at 7.5 years, and $3,100 at 11.5 years after the patent has been issued. In effect, a patent holder gets to enjoy the full twenty-year term only if he "renews" his patent three times. One study finds that 82.6 percent of patents were still in force (that is, had been "renewed") after four years, 57.4 percent after eight years, but only 37 percent after twelve years.[27] Using the same methodology as in Chapter 8, we have estimated depreciation rates from these data of 4.8 percent over the first four years after issue, 6.9 percent from five to eight years, and 8.3 percent thereafter. Over the entire twenty-year period the depreciation rate is about 6 percent. Put differently, we estimate an average economic life for a patent (given maintenance fees) of about 16.6 years, including a full twenty-year term for about 30 percent of issued patents.

25. See Adam B. Jaffe, Michael S. Fogarty, and Bruce A. Banks, "Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation," 46 *Journal of Industrial Economics* 183 (1998). The authors cite several previous studies of patent citations. Id. at 185. The authors tried to verify the accuracy of the citations, and found that 75 percent were meaningful, the rest essentially noise. Id. at 202. Laura M. Baird and Charles Oppenheim, "Do Citations Matter?" 20 *Journal of Information Science* 2, 7 (1994), estimates that at least 20 percent of patent citations are erroneous.

26. Many foreign countries grant renewable patents. For a summary of studies of such renewals, see Jean O. Lanjouw, Ariel Pakes, and Jonathan Putnam, "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data," 46 *Journal of Industrial Economics* 405 (1998); for an illustrative study, see Mark Schankerman and Ariel Pakes, "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period," 96 *Economic Journal* 1052 (1986).

27. Mark A. Lemley, "Rational Ignorance at the Patent Office," 95 *Northwestern University Law Review* 1495, 1503–1504 (2001) (1998 data).

Since the "renewal" fees are pretty stiff (certainly compared to comparable fees respecting copyrights and trademarks), our estimates of patent depreciation, combined with data on depreciation of copyrights, provide some basis for thinking that patents create substantial private values—but by the same token that they may confer greater monopoly power than copyrights do, as indeed we would expect given the nature of the two types of right. In Chapter 8 we estimated for the period 1934 to 1991 depreciation rates of 13.4 percent for graphic-arts copyrights, 9.2 percent for copyrights on books, 6.5 percent for trademarks, and 4.1 percent for copyrights on music, compared to our 6 percent estimate in this chapter for patents. The interesting comparison is between patents, trademarks, and music copyrights, on the one hand, and book and graphic-arts copyrights on the other. The depreciation rates are much lower in the first triad even though patent and trademark renewal fees are much higher than copyright renewal fees. We speculated that trademark depreciation rates were lower than those of most classes of copyright because a trademark can readily be transferred to another product or service; the trademark right thus has "breadth." The depreciation rate of music copyrights is low, we suggested, because music, often lacking words (and even when there are words, they are often an unimportant element of the piece) and susceptible of an indefinite number of different arrangements, is more adaptable to different tastes and times than books are, while graphic-arts works, with the occasional exception (such as Mickey Mouse), tend to be highly ephemeral because they usually are tied to particular advertising campaigns or product cycles. The low depreciation rate of patents is particularly striking because of the stiff fees; the combination of low fees and the high elasticity of renewals to fees persuaded us in Chapter 8 that most copyrights have little commercial value. Evidently patents are on average much more valuable than copyrights, and comparison with trademarks and music copyrights suggests that part of the reason is that they indeed cover more ground than copyrights do, other than music copyrights.

Do they cover more ground because otherwise the inventor could not recover his fixed costs? The question is empirical and the evidence inconclusive. Many highly progressive, research-intensive industries, notably including the computer software industry, do not rely heavily on patents as a method of preventing free riding on inventive activity.[28] For there are alternative ways of preventing free riding on valuable inventions. Recall from our initial discussion of copyright, in Chapter 2, that the likelihood that copying will actually prevent the creator of intellectual property from recovering his fixed costs depends on the cost of copying. That cost is often high in the case of industrial innovations because of the learning curve. If the cost of using some process or making some product decreases (up to a point) with time as the user becomes more skilled and experienced in the use of the process or the manufacture of the product, the imitator will find himself at a cost disadvantage in competing at the manufacturing stage. This disadvantage may offset the imitator's advantage in not having borne any of the costs of the innovation itself.

28. See, for example, Richard C. Levin et al., "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity* 783 (1987); Mark Schankerman, "How Valuable Is Patent Protection? Estimates by Technology Field," 29 *RAND Journal of Economics* 77 (1998); Robert Mazzoleni and Richard R. Nelson, "The Benefits and Costs of Strong Patent Protection: A Contribution to the Current Debate," 27 *Research Policy* 273, 275–276 (1998); Wesley M. Cohen, Richard R. Nelson, and John P. Walsh, "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)" (National Bureau of Economic Research Working Paper No. 7552, 2000).

May, but also may not. In the case of new drugs, whose manufacturers are avid in seeking patent protection,[29] the fixed costs of research and development are very high, in part because of stringent regulatory requirements, but marginal costs are very low, including the marginal costs of imitators. (The same thing is true of computer software, but trade secrecy and copyright law give software manufacturers alternative intellectual property protection to patents.) On a present-value after-tax basis, R&D is 30 percent of the total cost of a new drug,[30] although questions have been raised about the social productivity of much of this R&D.[31] It is also the case that some drug patents are unusually difficult to invent around,[32] which may make them too lucrative. Recall that, as illustrated by Figure 11.1, the more difficult it is for competitors to invent around a patent, the less elastic will be the patentee's residual demand curve and hence the higher will be the markup of price over marginal cost and so the greater the gross profits generated by the patent.

The difficulty of inventing around drug patents is not solely technological. There is evidence that when branded drugs, the patents on them having expired, must compete with much cheaper generic drugs (cheaper because they are free riders on the R&D of their brand-name predecessors), their prices

29. "In only one industry, drugs, were product patents regarded by a majority of respondents as strictly more effective than other means of appropriation." Levin et al., note 28 above, at 796.

30. Patricia M. Danzon, *Pharmaceutical Price Regulation: National Policies versus Global Interests* 5 (1997).

31. About two-thirds of drug R&D is academic and federal rather than industry, Darren E. Zinner, "Medical R&D at the Turn of the Millennium," *Health Affairs,* Sept./Oct. 2001, pp. 202, 205 (exh. 4), and it has been argued that most drug breakthroughs are due to academic and federal rather than industry research. See, for example, Public Citizen Congress Watch, "Rx R&D Myths: The Case against the Drug Industry's R&D 'Scare Card'" 8–10 (Washington, D.C., July 2001).

32. See Eric von Hippel, *The Sources of Innovation* 53 (1988); Levin et al., note 28 above, at 798.

tend not to fall.[33] As we noted in Chapter 7, the patent monopoly accustoms doctors and patients to the name-brand product (trademark reinforcing patent), and when the patent expires and a generic substitute becomes available at much lower cost, they remain reluctant to substitute for the familiar brand an unknown (un)brand, which though certified as chemically identical may differ in some subtle way, perhaps involving quality control. The sales of the branded drug will fall, as the manufacturer in effect cedes the low-price segment of the market to the generics, but profit per unit will remain high, and aggregate profits, though smaller than before, will remain healthy. One study found only a gradual decline in market share upon patent expiration,[34] though analysis of the welfare effects is complicated by the possibility that a patentee, in order to perpetuate its monopoly, will invest in brand loyalty by reducing the price of the patented product in the pre-expiration period.[35]

Much of the evidence relating to drug patents predates, however, the passage of the Hatch-Waxman Act in 1984. The Act allows generic-drug manufacturers to begin the required FDA testing of their drug before the patent on the brand-name equivalent expires, without being guilty of patent infringement. Apparently the Act, along with other initiatives private as well as public to control spiraling costs of health care, has produced a large increase in generics' market share.[36] However, the Act also extended the patent term for pharmaceutical drugs by the amount of time (up to a maximum of five

33. See, for example, Roger D. Blair and Thomas F. Cotter, "Are Settlements of Patent Disputes Illegal Per Se?" 47 *Antitrust Bulletin* 491, 496–501 (2002); Dong-Churl Suh et al., "Effect of Multiple-Source Entry on Price Competition after Patent Expiration in the Pharmaceutical Industry," 35 *Health Services Research* 529 (2000); F. M. Scherer, "Pricing, Profits and Technological Progress in the Pharmaceutical Industry," *Journal of Economic Perspectives,* Summer 1993, pp. 97, 101–102. Furthermore, if the patentee obtains a follow-on patent during the patent term and claims that the generic will infringe it, the FDA will automatically grant repeated thirty-month stays of approval of the generic until the patent dispute is resolved. See, for example, Robert Langreth and Victoria Murphy, "Perennial Patents," *Forbes,* Apr. 2, 2001, p. 52. This practice is under attack and may soon be changed by regulation or legislation.

34. See Meir Statman, "The Effect of Patent Expiration on the Market Position of Drugs," 2 *Managerial and Decision Economics* 61 (1981). There is even evidence that branded-drug prices rise when patent protection ceases because the manufacturers, writing off the high-elasticity segment of the market, increase their marketing efforts in the low-elasticity segment. See Ernst R. Berndt, "Pharmaceuticals in U.S. Health Care: Determinants of Quantity and Price," *Journal of Economic Perspectives,* Fall 2002, pp. 45, 63; Steven J. Davis, Kevin M. Murphy, and Robert H. Topel, "Entry, Pricing and Product Design in an Initially Monopolized Market" (National Bureau of Economic Research Working Paper No. 8547, Oct. 2001).

35. See Gideon Parchomovsky and Peter Siegelman, "Towards an Integrated Theory of Intellectual Property," 88 *Virginia Law Review* 1455 (2002).

36. See Berndt, note 34 above, at 62–63; Henry Grabowski, "Patents and New Product Development in the Pharmaceutical and Biotechnology Industries" 6–8 (Duke University Department of Economics, July 2002). See also Jay P. Bae, "Drug Patent Expirations and the Speed of Generic Entry," 32 *Health Services Research* 87 (1997).

years) that it takes for the drug to be approved by the Food and Drug Administration for sale to the consuming public.

In creating a testing exception to infringement, the Hatch-Waxman Act expanded, though only in the pharmaceutical domain, the long-standing but narrowly interpreted "experimental use" doctrine of patent law, the counterpart to the fair use doctrine of copyright law. It has been urged that the experimental use doctrine be expanded to allow scientists to use patented research tools (such as gene fragments, and the tumor-prone oncomouse used, as its name implies, in cancer research) without license. The main argument is that the number of patented research tools required to conduct experiments is often so great that the transaction costs of obtaining licenses for all of them are prohibitive.[37] The doubts that we have expressed concerning the social benefits of the existing level of patent protection argue for generous construal of fair use principles in patent law as in copyright law.

That the patent term for new drugs may be too long is not refuted by evidence that, contrary to widespread belief, the profits of the U.S. pharmaceutical industry do not significantly exceed the industry's cost of capital.[38] The evidence is consistent with government regulation that limits the ability of drug manufacturers to charge monopoly prices to certain segments of the population. It also suggests that the manufacturers of differentiated drugs are competing with each other—an example of monopolistic competition, the situation in which each seller (which might be as humble as a barber shop enjoying a slight locational monopoly) in a market faces a downward-sloping demand curve but because entry is unimpeded, the firms in the market have zero monopoly profits, price being equal to average cost.[39] Competition for monopoly rents will, as we know, tend to transform them into costs without necessarily producing commensurate social benefits. In the example in Chapter 1 of a race to be the first to discover and appropriate a new continent, the ten competitors as a group had zero profits, but because they were competing for rents their costs were higher than otherwise and exceeded the benefit, generated by the higher costs, of discovering the continent a year earlier. For the same reason, the fact that much drug research fails to generate products that recoup the cost of the research is consistent with rent seeking. The prospect of large profits enabled by patent protection provides a lure for investment in research, yet the resources devoted to that research conceivably

37. See Janice M. Mueller, "No 'Dilletante Affair': Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools," 76 *Washington Law Review* 1 (2001).

38. See Henry G. Grabowski and John M. Vernon, "Returns to R & D on New Drug Introductions in the 1980s," 13 *Journal of Health Economics* 383 (1994). See also Scherer, note 33 above, at 103–106.

39. See Carlton and Perloff, note 7 above, at 454–457.

might be socially more productive in an industry in which innovation is not rewarded with a monopoly.

Nevertheless the strongest case for patents in something like their present form is said to be found in a subset of the drug industry: "The collection of small and medium sized firms in the American biotechnology industry is, of course, a striking example of enterprises that would not have come into existence without the prospect of a patent, and which depend on patent protection to make their profits, and to attract capital."[40] Yet there is concern that the extent of patent protection of biological research tools may be such as to impede biotechnological progress. For example, the existence of separate patents on complementary gene fragments may make the transaction costs of assembling genetic material needed for research very high. More generally, the licensing of research tools is complicated and therefore costly for reasons similar to why patent licensing of basic research would be complicated and therefore costly.[41] At the same time, the industry has a valid concern that a "fair use" (or broadened experimental use) exception for research tools developed and patented by industry would enable academic researchers to use the tools—and thus free ride on the industry R&D—to create and obtain their own patents on new drugs.

A natural experiment on the effects of patents is the federal government's encouragement of applied research by universities, which has resulted in a great increase in university patenting. But the implications for industry are unclear. Universities are unusual in being engaged in research but not production. It was natural that they would focus on basic research, and it would have been difficult for them to profit from conducting applied research if the fruits could not be patented, since they could not profit by embodying the research in products or processes that they could insulate from competition by trade secrecy or first-mover advantages, but only by licensing their inventions. Being able to earn substantial income from patent licensing has, it appears, induced universities to substitute away from basic research,[42] and the result may have been a net social loss.

The difficulty of evaluating the social benefits of the patent system can be

40. Mazzoleni and Nelson, note 28 above, at 276. Nelson is a leading economist-skeptic about patents, so the quoted statement carries particular weight.

41. See *Report of the National Institutes of Health (NIH) Working Group on Research Tools,* June 4, 1998, *www.nih.gov/news/researchtools/index.htm;* Michael A. Heller and Rebecca S. Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," 280 *Science* 698 (1998). But see the contrary arguments, reflecting the industry view, in Richard A. Epstein, "Steady the Course: Property Rights in Genetic Material" (University of Chicago Law School, John M. Olin Law and Economics Working Paper No. 152 [2d ser.], June 2002).

42. See note 24 above.

illustrated by reference to another legal limitation on patents as compared to copyrights. An inventor can obtain a patent on an improvement that he makes to someone else's patent—provided the improvement satisfies the normal requirements for patentability, including utility, novelty, and nonobviousness—without obtaining a license from the original patentee. It's as if the owner of a derivative work could copyright it without a license from the owner of the copyright on the original. The patentee of the improvement probably won't be able to use his patent without infringing the original patent, but by the same token the original patentee cannot make the improvement without infringing the improver's patent. This is the situation of blocking patents; it forces a negotiation between the patentees if the most efficient technology is to be employed.[43] Whether the extra negotiation costs offset the potential incentive benefits from encouraging independent firms to develop improvement patents is unclear. A full analysis would also require consideration of the possible diminished incentives of the original patentee to develop an improvement patent, and the saving in licensing and search costs if the patentee had these rights to begin with and sought out other firms to improve his patent.

Awareness of the problem of bilateral monopoly, and more broadly of transaction costs, in the patent setting explains the "reverse doctrine of equivalents."[44] Under that doctrine, if the contribution made by the improvement *greatly* exceeds the contribution made by the original patented invention, the improver is allowed to practice his invention without being deemed an infringer, even though he is making use of the prior invention without a license from the patentee. Under the "doctrine of equivalents," discussed below—a secure part of patent law—small differences between a patented invention and an alleged infringing invention are not a defense to infringement. The rationale of the reverse doctrine is that requiring the improver to negotiate a license from the original inventor would impede a potentially valuable improvement; this is a transaction-costs rationale. It reflects fair use thinking transposed from copyright to patent law: when the improver makes only a trivial use of the patented invention, transaction costs swamp the social benefit of allowing the patentee to exact a licensing fee.

This raises the question whether copyright law should emulate patent law's

43. See Westinghouse v. Boyden Power Brake Co., 170 U.S. 537 (1898); Robert P. Merges and Richard R. Nelson, "On the Complex Economics of Patent Scope," 90 *Columbia Law Review* 839 (1990); Howard F. Chang, "Patent Scope, Antitrust Policy, and Cumulative Innovation," 26 *RAND Journal of Economics* 34 (1995).

44. See, for example, Merges and Nelson, note 43 above. The doctrine has received a frosty reception from the Federal Circuit, but this may be changing. See Amgen Inc. v. Hoechst Marion Roussel, Inc., 314 F.3d 1313, 1351 (Fed. Cir. 2003).

treatment of improvements. There is a clue to the answer in the rules on co-ownership. Patent law allows any co-owner of a patent to use it without the permission of the other co-owners. This is similar to the rule of copyright law that any of the coauthors of a joint work may exercise the full rights of a copyright holder (to use the work, license its use to others, etc.), subject to having to share the profits from the license or other use with his coauthors (share equally, in the absence of an agreement specifying the parties' respective shares). Both are rules designed to minimize bilateral monopoly. The patent rule differs from the copyright rule, however, in not requiring an accounting for profits to the co-owners (unless they agree to share the profits). The reason may be that technological advance is a continuous process of improvement, and transaction costs are minimized if a patentee can work on the patented invention or license others to do so without sharing the profits with his co-owners. Continuous improvement is less common in expressive works, though of course not unknown—musical arrangements, cartoon characters, and of course computer software are examples of copyrightable works that often undergo a process of continuous development.

Even though transaction costs would be minimized in one sense if patent improvements were treated like derivative works in copyright law, they would be increased in another sense if one assumes that the original patentee and his licensees (if any) cannot always be depended on to recognize opportunities for improvements to his invention. Many students of innovation believe that it is best understood as a quasi-Darwinian process—a process almost of trial and error in which the market selects from among diverse approaches whose relative promise cannot be assessed in advance.[45] This approach implies that a multiplicity of independent sources of inventive activity is superior to a centralized process directed by the patentee. Although the patentee has an incentive to license inventors of improvements in its patents, its ability to act on that incentive may be impeded by firm culture, management style, the hierarchic, bureaucratic structure of a large firm, the quirks of particular employees, and other factors that may be difficult to control and that vary from firm to firm.[46] There is only so much diversity that an organization can tolerate, as we learn from the frequent failure of mergers because the merging firms turn out to be unable to meld their distinct cultures. Intrafirm diversity in inventing or in licensing inventors is thus an imperfect substitute for interfirm diversity. The Darwinian theories of invention provide economic support for the improvement patent,[47] which encourages firms to do R&D on technologies dominated by patents held by other firms.

45. See, for example, Richard R. Nelson and Sidney G. Winter, *An Evolutionary Theory of Economic Change* (1982); Nelson and Winter, "Evolutionary Theorizing in Economics," *Journal of Economic Perspectives,* Spring 2002, pp. 23, 33–39; Merges and Nelson, note 43 above, at 873–879.

46. "Firms facing the same market signals respond differently, and more so if the signals are relatively novel." Nelson and Winter, *An Evolutionary Theory of Economic Change,* note 45 above, at 276.

The same approach could be taken to copyrights, and derivative works thus made copyrightable without license from the owner of the copyright on the original work.[48] We have acknowledged in this chapter that some copyrighted works undergo continuous improvement, and we acknowledged in Chapter 4 that allowing unauthorized derivative works to be copyrighted might lead to increased creativity. But there is considerable potential for confusion in such an approach because of the low cost of making many types of derivative work and the fact that they are not screened in advance for originality, or held to a high standard of originality if challenged in court, as patents are, and copyrighting costs literally nothing because copyright attaches automatically to a copyrightable work. The mischief potential of "blocking" copyrights would therefore be high. In contrast, it is costly to obtain a patent (apart from the major cost, that of R&D, it costs about $10,000 to $30,000 in filing fees, attorneys' fees, and other expenses to prepare a patent application and get it approved by the Patent and Trademark Office),[49] including a patent on an improvement. Money costs to one side, because trivial improvements will flunk the criteria for patentability, the danger that optimal development of a patented technology will be blocked (in the blocking-patents sense) by trivial improvements, resulting in high transaction costs without offsetting benefits in enhanced innovation, is minimized.

The Darwinian theories of innovation have a further significance for analysis of the patent system. They cast doubt on Edmund Kitch's prospect theory,[50] which commends the patent system for centralizing the inventive process in the original "prospector." The original prospector may have a flawed conception of the optimal path of development. If as Kitch believes it is the ***reduction*** in the amount of duplication of inventive activities that is the prime merit of the patent system, the system may, while effecting some economies, actually be retarding technological progress. Another objection to Kitch's theory is that while its objective is to reduce rent seeking, specifically patent races, by increasing the benefits to being the original prospector, the theory if

47. See Merges and Nelson, note 43 above, at 873–879; Mark A. Lemley, "The Economics of Improvement in Intellectual Property Law," 75 *Texas Law Review* 989 (1997).

48. As argued by Professor Lemley in the article cited in the preceding note.

49. See Lemley, note 27 above, at 1498. And the cost in attorneys' fees is likely to be higher if it is a valuable patent and the owner therefore wants to maximize the likelihood that it will withstand a court challenge.

50. See Merges and Nelson, note 43 above, at 872–875.

implemented might engender wasteful races to be that prospector. The earlier the patent is granted, the broader its protection is likely to extend (because there will be less prior art to be skirted by narrowing the claims in the patent application)—making it more valuable and inciting a greater expenditure on trying to be first to obtain the patent.

Furthermore, patents often are sought not because the applicant considers patenting a more effective method of recapturing his fixed costs of innovation than trade secrecy or lead time (his head start over competitors and the resulting learning-curve advantage that will persist after his competitors imitate him), but because he wants to prevent others from obtaining a patent that might be used to prevent him from using his innovation without paying someone else a licensing fee.[51] The more readily patents are granted and are upheld in court and the broader the legal protection they confer, the greater the incentive for defensive patenting of the kind just described, patenting not motivated by inability to recover the fixed costs of invention by other means.

Defensive patenting must be distinguished from patent suppression.[52] The

51. See Adam B. Jaffe, "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process," 29 *Research Policy* 531, 539–540 (2000); John R. Allison and Mark A. Lemley, "Who's Patenting What? An Empirical Exploration of Patent Prosecution," 53 *Vanderbilt Law Review* 2099, 2104 n. 17 (2000); Nancy T. Gallini, "The Economics of Patents: Lessons from Recent U.S. Patent Reform," *Journal of Economic Perspectives*, Spring 2002, pp. 131, 140, 149; Levin et al., note 28 above, at 798 n. 29. An alternative is simply to publish one's invention in the hope that this will convince the PTO to turn down any application to patent it on the ground that it lacks novelty. See Douglas Lichtman, Scott Baker, and Kate Kraus, "Strategic Disclosure in the Patent System," 53 *Vanderbilt Law Review* 2175 (2000).

52. A recent statement by a patent attorney for Hewlett Packard, though it uses language of suppression, appears to be referring, rather, to defensive patenting: "'We get patents not to protect our own products, but because it gives us power to exclude in areas where others might want to participate,' he recently told about a dozen H-P researchers and scientists gathered near the company's headquarters here. 'We assume our competitors are filing for patents in all different areas. We don't want to be the last ones on the block.'" Pui-Wing Tam, "More Patents, Please! Tech Companies Urge Staffers to Submit Innovative Ideas; Cash Awards, Plaques at H-P," *Wall Street Journal*, Oct. 3, 2002, p. B1. For a more explicit description of the defensive strategy, see Russell L. Parr, "IP Leverage: Facilitating Corporate Value Creation," in *From Ideas to Assets: Investing Wisely in Intellectual Property* 271, 282 (Bruce Berman ed. 2002): "A defensive strategy is simple. Patent everything in sight and threaten competitors with infringement litigation when they come too close to making products or doing business in a similar fashion. Licensing income is not a goal that is part of this strategy." The importance of defensive patenting is suggested (no stronger word is possible) by the estimate that "at any given time, over about 95 percent of patents are unlicensed and over about 97 percent are generating no royalties." Samson Vermont, "The Economics of Patent Litigation," in id. at 327, 332. Cf. Bronwyn H. Hall and Rosemarie Ham Ziedonis, "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979–1995," 32 *RAND Journal of Economics* 101, 125 (2001). The role of patents in facilitating cartelization is an old story, but a true one; we discuss it in Chapter 14.

defensive patentee is not trying to prevent the emergence of a new technology. But there are a number of well-documented cases of firms' acquiring or developing a new technology, patenting it, and then deciding not to make or license the patented product even though it is commercially promising.[53] Such patent suppression can be economically rational behavior.[54] Suppose *A* and *B* are competitors, and *A* believes that there is a new technology that would be compatible with *B*'s production methods but not with his own and that if adopted by *B* would give *B* a decisive competitive advantage. *A* might in these circumstances have a rational incentive to expend some resources on preventing *B* from adopting the new technology, by patenting it first. An alternative would be to license the use of the patent to *B* at a royalty rate high enough to extract most of the benefit of the technology to *B*. But *A* might be reluctant to do that because of fear that in working with the new technology *B* would develop still better technology and so steal a march on *A*. It might be costly to calculate a license fee that would protect *A* from being harmed by such an eventuality—or to negotiate the fee without revealing *A*'s concern to *B*.

Probably the most common reason for patent suppression, however, is innocent: after obtaining the patent the patentee gets cold feet, doubting that the expense of actually producing the new product or adopting the new process is commensurate with the expected return. Licensing remains an option but involves, as just suggested, significant transaction costs.

We need not delve any deeper into the issue of patent suppression.[55] Our point is only that patents are sometimes suppressed, and this is an example of how they can actually impede technological progress.

Defensive and suppressive patenting may be grouped together as prime examples of "strategic" patenting, the subject of a substantial business literature that casts additional doubt on the efficiency of the patent system as a means of optimizing the rate and direction of inventive activity. In that literature we come across frequent remarks of the sort that "invention is not the point of most valuable patents. Instead, most patents are obtained for the proper business purpose of keeping competitors away from the market for a new product

53. See the exhaustive treatment in Kurt M. Saunders, "Patent Nonuse and the Role of Public Interest as a Deterrent to Technology Suppression," 15 *Harvard Journal of Law and Technology* 389 (2002).

54. See Carlton and Perloff, note 7 above, at 538; Paul Stoneman, *The Economic Analysis of Technology Policy* 113–114 (1987).

55. Thus we need not consider the interesting suggestion in Julie S. Turner, Comment, "The Nonmanufacturing Patent Owner: Toward a Theory of Efficient Infringement," 86 *California Law Review* 179 (1998), that a patent owner who does not intend to use his patent should not be entitled to enjoin infringers, but instead should be remitted to his remedy in damages—which would be slight in most cases, at least if monopoly profits generated by patent suppression were not thought to be the kind of loss for which patent law should authorize damages.

or service."[56] "The opposite of building a patent wall around your product—or clustering, as it is sometimes called—is bracketing the patents of your competitors."[57] "Companies have been said to use a technique of patent 'flooding' or 'blanketing' a technology area . . . The typical scenario is that a new technology is patented by a first company, and a second company . . . if the stakes are high enough, can assign enough resources literally to blanket all of the potential improvements to the invention by filing patents on these improvements. The first company is then essentially forced into some type of cross-licensing agreement if they want a business to grow."[58] There may well be exaggeration here; aggressive rhetoric comes easily to business executives and consultants and often has no economic significance. And the improvements patented by these "bracketers" or "blanketers" are not worthless—if they were not real improvements the original inventor would have no incentive to seek to license them. Yet it might be more efficient to leave the improvements to be made, at a slower pace but at a lower cost, by the original inventor.

Mention of improvement patents brings into view the question of what shall count as an infringement, a question that further and deeply complicates the assessment of the patent system's overall economic effects. We pointed out in Chapter 4 that from an economic standpoint whether a new expressive work infringes a copyrighted work should depend on whether the extent of the copying of the old work by the new is sufficient to make the new a close substitute for the old in the marketplace. The same is true regarding patents. But substitution is a matter of degree, and it is unclear whether patent grants should be interpreted broadly, by an expansive construal of the "doctrine of [patent] equivalents,"[59] or narrowly. The doctrine of equivalents corresponds to the concept of substantial similarity used to determine whether a copy of an expressive work is close enough to infringe the copyright on that work. It is nicely illustrated by *International Nickel Co. v. Ford Motor Co.*[60] International Nickel had a patent on "modular iron," which involved a .04 percent addition of magnesium to molten iron. Ford began making its own version of modular iron, which solely differed from International Nickel's in containing only .02 percent magnesium. The court held that Ford had infringed International Nickel's patent; Ford's variant was "equivalent" to the patented product. Had Ford made a substantial improvement and otherwise satisfied

56. Stephen C. Glazier, *Patent Strategies for Business* 11 (3d ed. 2000).

57. Kevin J. Rivette and David Kline, *Rembrandts in the Attic: Unlocking the Hidden Value of Patents* 110 (2000).

58. H. Jackson Knight, *Patent Strategy for Researchers and Research Managers* 42–43 (1996).

59. Reaffirmed recently in Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 535 U.S. 722 (2002).

60. 166 F. Supp. 551 (S.D.N.Y. 1958), discussed in Merges and Nelson, note 43 above, at 853–854.

the requirements of patentability, it could have obtained an improvement patent.

The practical test for equivalence requires supposing that the patentee's patent application had contained a claim for the very invention that he now claims is equivalent to the patent that he did obtain, and then asking whether such a claim would have been barred by a previous patent or other prior art.[61] This is similar to asking in a copyright case whether the copyright holder is seeking protection for expression or for an idea. If the latter—if for example he is seeking to prevent someone else from using some standard plot on the ground that (in patent jargon) the defendant's use of that plot makes the defendant's work "equivalent" to his own though not identical to it—the judicial response would be that he could not have copyrighted the element that he claims makes the defendant's work equivalent to his.

The doctrine of equivalents enables patent applicants to economize on description. It thus invites comparison to doctrines of contract law that by supplying standard terms economize on the costs of drafting contracts. Without a doctrine of equivalents, patent applicants would have to include much greater detail concerning the extent of the patent protection sought. The doctrine is susceptible, however, to the following abuse: apply for as broad a patent as possible, that is, be as abstract as possible. (Recall, again from Chapter 4, the correlation between breadth and abstractness.) If the broad patent is granted, you may have gotten yourself a very substantial monopoly. Samuel Morse, the inventor of the telegraph, applied for a patent on "the use of the motive power of . . . electro-magnetism, however developed[,] for making or printing intelligible characters at any distance." A patent so broad might have returned Morse profits vastly in excess of his fixed costs of invention, especially if he could in effect have extended his seventeen-year patent term by patenting improvements made during the term.[62] If the PTO balks at the breadth of the patent protection sought, the applicant can amend the application to narrow the scope—and then after the application is granted reassert the abandoned claims under the doctrine of equivalents (that is, argue that any patent containing any of those claims is equivalent to his own patent). This maneuver is blocked by the doctrine of "prosecution history estoppel," which as recently interpreted by the Supreme Court presumptively forbids the patentee to invoke the doctrine of equivalents with respect to any claim that the PTO, in the course of insisting on a narrower application, rejected as failing to satisfy one or more of the requirements for patentability.[63] Another check

61. See Wilson Sporting Goods Co. v. David Geoffrey & Associates, 904 F.2d 677 (Fed. Cir. 1990).

62. See O'Reilly v. Morse, 56 U.S. (15 How.) 62 (1853), discussed in Merges and Nelson, note 43 above, at 850–851.

63. In the *Festo* case, note 59 above.

on overclaiming in a patent is the statutory rule that a patent claim is invalid if it is indefinite; an indefinite claim would tend to enlarge the practical scope of the patent beyond its lawful bounds by imposing legal risks on competitors, who would be buying an infringement suit if they mistook those bounds because the patent claim was unclear. The requirement would be less important if there were a defense of independent creation, as in copyright law; since there is not, the potential infringer is entitled to a clear warning of the scope of the patentee's rights, so that he does not infringe inadvertently.

Broad interpretation of a patent's scope increases the patentee's power to exclude competition. But at the same it "forces other firms, if they want to compete in the broad product field, to work on alternatives that may be very different from what is already patented."[64] It may also (shades of Kitch) reduce patent racing,[65] though, alternatively, it may just shift the patent race to an earlier period (the race for the broad patent). Narrow interpretation can increase transaction costs: subsequent inventors may have to get licenses from more patentees[66]—though the other side of the coin is that the narrower the existing patents are, the less likely a new invention is to infringe any of them. It has been argued that narrow interpretation may, in addition, actually increase the deadweight loss from the patentee's monopoly pricing by making it easier to create close substitutes that cost considerably more to make than the patentee's product.[67] Suppose the marginal cost of that product is $1 and the price $3, but if the patent grant is construed narrowly a competitor can make a close substitute at a cost of $2. Consumers will be deflected to the competitor. The patentee may respond by reducing his price in an effort to win them back. Whether this process enhances social welfare is uncertain. On the one hand, deadweight losses from output restrictions are reduced; on the other hand, production costs are higher when both the original patentee and a competitor or competitors produce two products than when a product is produced by the patent holder alone. Another consideration is that to the extent the competitor is successful, the patent reward is reduced and the patentee may be unable to recoup his fixed costs of invention. If instead, as a consequence of a broad interpretation of the patent, the substitute

64. Mazzoleni and Nelson, note 28 above, at 275.

65. See, illustrative of a large literature, Vincenzo Denicolò, "Patent Races and Optimal Patent Breadth and Length," 44 *Journal of Industrial Economics* 249 (1996). We do not discuss the tradeoff between breadth and duration of patents; in effect, we treat the existing duration (generally twenty years) as exogenous to our analysis.

66. See Robert P. Merges, "Institutions for Intellectual Property Transactions: The Case of Patent Pools," in *Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society* 123 (Rochelle Cooper Dreyfuss, Diane Leenheer Zimmerman, and Harry First eds. 2001).

67. See Paul Klemperer, "How Broad Should the Scope of Patent Protection Be?" 21 *RAND Journal of Economics* 113 (1990).

is deemed infringing, fewer consumers will substitute away from the patented product because they will be denied access to the closest substitute.[68]

But the analysis is incomplete. A broad patent, by greatly limiting substitution, will enable the patentee to charge a higher price than if the patent were narrow.[69] Because a profit-maximizing monopolist will always raise his price to a point within the elastic range of his demand curve, the higher price will deflect some consumers to substitutes, albeit poor ones, offsetting (to what extent cannot be determined a priori) the effect of a broad patent in compressing the range of available substitutes.

Here is another equivocal benefit of broad interpretation: in increasing the frequency of defensive patenting, it increases the number of patents, and the more patents that are issued, the more disclosure of technological ideas is made that other inventors may be able to use.[70] But if defensive patenting is a major factor in the increase in the number of patents in the last two decades (see the statistics in Chapter 12), and thus an indirect result of the strengthening of patent protection over this period that may underlie the increase, further doubt is cast on whether more patents mean more and better innovation.

Broad patent protection has still another, and fundamental, double-edged effect: it increases the return to the first inventor, which encourages invention, but increases the cost of invention to his successors, which discourages invention. The analogy to our discussion of copyright in Chapters 2 and 3, where we pointed out that the broader copyright protection is, the costlier the subsequent creation of expressive works becomes because earlier works are inputs into later ones, should be apparent. A related point is that the more exacting the requirement of novelty and so the harder it is to get a patent, the easier it is for subsequent inventors to obtain patents (although they will be narrow too) but the less information useful to them will be disclosed in patent applications because there will be fewer of them.[71]

The analogy of patent to copyright invites an adaptation of the formal model in Chapter 3 to patents, going beyond Figure 11.1 and the accompanying discussion. Indeed, at the formal level, the parallelism is so close that we could use the identical model, deriving the identical results, with just a

68. A similar argument was made with respect to copyright in Ian E. Novos and Michael Waldman, "The Effects of Increased Copyright Protection: An Analytic Approach," 92 *Journal of Political Economy* 236, 244–245 (1984). Recall that our own doubts about broad copyright protection are not based on underutilization (that is, reduced access by consumers), but on transaction costs and input costs, which the Novos and Waldman article does not discuss.

69. We ignore the complications introduced by price discrimination; they would not alter the analysis fundamentally.

70. See Gallini, note 51 above, at 140.

71. See Suzanne Scotchmer and Jerry Green, "Novelty and Disclosure in Patent Law," 21 *RAND Journal of Economics* 131 (1990). A further argument for narrow interpretation of patents is presented in the next chapter.

relabeling of the variables: p would be the price of a unit of output of the patented product, $q(p)$ the market demand for that output, x and $y(p, z)$ the number of units produced by the patentee and the duplicators (where y depends on both p and z, the amount of protection against duplication conferred by patent law), c the patentee's marginal cost, e the cost of creating the patented invention in the first place, and so forth. The implications for patent law and inventions would with minor adjustments be similar to those listed at the end of Chapter 3 for copyright law and expressive works.

But wouldn't there have to be a major adjustment to reflect the boon that patents confer on subsequent patentees by forcing disclosure of the steps constituting the invention? This might seem to alleviate the tension between earlier and later innovators that we emphasized in the copyright setting. But the appearance is misleading. For if we set unpublished materials and some computer software[72] to one side for the moment, copyright law evokes disclosure just as patent law does. With the exceptions just mentioned, copyrighted works are fully public and subsequent authors can use any of the information (that is, ideas) contained in them to help them in making their own such works. There is even a term for this use—"managed copying." The differences between patent and copyright law, and between expression and invention, are great, but not at the level of abstraction at which our formal model was pitched.

Patent Law as a Response to Trade Secret Law and Monopoly

The foregoing analysis suggests grounds for skepticism that the existing level of patent protection is essential to enabling inventors to recover their fixed costs. These grounds are reinforced by a growing body of empirical studies illustrated by a study of Japanese patent law that, using patent citations and other proxies for evaluating the relative contribution of different patents to technological progress, found that Japan's expansion of patent rights in 1988 had no effect on innovation or R&D.[73] There is also evidence that the patenting of computer software actually retards innovation because most software

72. Object code (machine code), as distinct from source code, is not readable by a human being or otherwise usable by him unless he has access to the source code. If, therefore, a software manufacturer copyrights only the object code, retaining the source code as a trade secret, public access to the copyrighted code will not generate information that other software writers can use.

73. See Mariko Sakakibara and Lee Branstetter, "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms," 32 *RAND Journal of Economics* 77 (2001). Other studies are cited and summarized in id. at 98–99. Mazzoleni and Nelson, note 28 above, summarize the arguments for doubting that the expansion of patent protection in recent decades has promoted a net increase in economic welfare. And Jaffe, note 51 above, at 555, after a careful survey of the economic literature concerning the expansion, concludes: "There is widespread unease that the costs of stronger patent protection may exceed the benefits. Both theoretical and, to a lesser extent, empirical research suggest this possibility."

innovation both builds on and complements existing software. Without the retardation introduced by patenting and the resulting need to negotiate licenses, software manufacturers would innovate more rapidly and each would benefit from the others' innovations, which, because of the sequential and complementary nature of the innovations in this industry, would enhance the value of the existing products.[74]

We do not question the importance of technological progress to economic welfare, the relation of R&D expenditures to such progress, or that because of positive externalities the social return to R&D exceeds the private return.[75] The issue is the effect of the actual patent system that we have, and of feasible variants that can be imagined, on R&D, weighting quantity of R&D by quality. The positive correlation between the strength of a nation's intellectual property laws and the nation's economic growth, capital-labor ratio, per capita income, and government funding of R&D[76] similarly does not establish causal relations.

One study found that "while the aggregate value of patent rights appears to be quite high, it is estimated to be only on the order of 10 to 15 percent of total national expenditures on R&D. Hence it is unlikely to be the major factor in determining the overall level of expenditures."[77] If this is right, then incremental increases in patent protection are unlikely to influence inventive activity significantly and incremental reductions might actually enhance economic welfare. Notice in this regard that any increase in patent protection, to the extent it succeeds in its objective of inducing additional inventive activity, creates additional competition, since patented products are often substitutes for other patented products. The additional competition will reduce the profitability of inventive activity, and hence the resources devoted to it, blunting the effect of the enhanced patent protection in preventing the manufacture of close substitutes for patented products and thus the incentive effect of the enhanced protection.[78] There is support for this proposition in the

74. See James Bessen and Eric Maskin, "Sequential Innovation, Patents, and Imitation" (MIT Sloan School of Management and Princeton University, July 2002).

75. See Charles I. Jones and John C. Williams, "Measuring the Social Return to R&D," 113 *Quarterly Journal of Economics* 1119 (1998).

76. See Dennis W. Carlton, "A Critical Assessment of the Role of Imperfect Competition in Macroeconomics," in *Market Behaviour and Macroeconomic Modelling* 73, 84–86 (Steven Brakman, Hans von Ees, and Simon K. Kuipers eds. 1998).

77. Zvi Griliches, Ariel Pakes, and Bronwyn H. Hall, "The Value of Patents as Indicators of Inventive Activity," in *Economic Policy and Technical Performance* 97, 120 (Partha Dasgupta and Paul Stoneman eds. 1987).

78. See Tomas Philipson and Frank R. Lichtenberg, "The Dual Effects of Intellectual Property Regulations: Within- and Between-Patent Competition in the US Pharmaceuticals Industry" (University of Chicago, George J. Stigler Center for the Study of the Economy and the State, Working Paper No. 178, Oct. 12, 2002). This is a point that we made in Chapter 4 with reference to increased copyright protection.