Abstract

This chapter surveys and synthesizes the deepening and widening theoretical landscape of intellectual property. Not surprisingly, the principal philosophical theory applied to the protection of utilitarian works - that is, technological inventions - has been utilitarianism. Utilitarian theorists generally endorse the creation of intellectual property rights as an appropriate means to foster innovation. Non-utilitarian theorists emphasize creators' moral rights to control their work. Many of these scholars draw upon multiple philosophical strands in constructing their analyses.

JEL classification: K11
Keywords: Copyrights, Patents, Utilitarianism

1. Introduction

The theory of intellectual property has not, until recently, attracted much philosophical interest or been the subject of deep controversy. Utilitarian theorists generally endorsed the creation of intellectual property rights as an appropriate means to foster innovation, subject to the caveat that such rights are limited in duration so as to balance the social welfare loss of monopoly exploitation. Non-utilitarian theorists emphasized creators’ moral rights to control their work. With the increasing importance of intellectual property in society and the development of particular new technologies, most notably digital technology and the decoding of genetic structure, the theory of intellectual property has attracted heightened interest. Economists and policy analysts have greatly enriched our understanding of the complex relationship between intellectual property protection and innovation and diffusion of technological advances. Non-utilitarian theories of intellectual property have proliferated in recent years, as philosophers and legal scholars have applied traditional and novel philosophical perspectives to the realm of intellectual property. This article surveys and synthesizes the deepening and widening theoretical landscape of intellectual property. While much of the
discussion transcends the law of any particular nation, the statutory and doctrinal examples are drawn principally from the particularities of the United States intellectual property regimes.

A. Utilitarian/Economic Theories of Intellectual Property

Not surprisingly, the principal philosophical theory applied to the protection of utilitarian works - that is, technological inventions - has been utilitarianism (Merges, et al., 1997, pp. 135-136 hereinafter cited as MMLJ; Machlup, 1958). The social value of utilitarian works lies principally if not exclusively in their ability to perform tasks (for example, a better mousetrap) or satisfy desires more effectively or at lower costs. It is logical, therefore, that society would seek to protect such works within a governance regime that itself is based upon utilitarian precepts. Furthermore, inventions - new processes, machines, manufactures, or compositions of matter - unlike artistic or literary expression do not generally implicate personal interests of the creator. (For a discussion of the application of non-utilitarian theories to patent law, see Oddi, 1996, pp. 274-277, discussing reward-based and natural law theories; Becker, 1993, noting intuitive appeal of entitlement-based arguments.) The United States Constitution expressly conditions the grant of power to Congress to create patent and copyright laws upon a utilitarian foundation: ‘to Promote the Progress of Science and useful Arts’. Economic theory, a particular instantiation of utilitarianism, has provided the principal framework for analyzing intellectual property.

In addition, the utilitarian perspective has relevance to other forms of intellectual property. Trade secret law often protects utilitarian works (MMLJ, 1997, pp. 34-36; Scheppele, 1988). Trademark law is principally concerned with ensuring that consumers are not misled in the marketplace and hence is particularly amenable to economic analysis (Economides, 1988). Even copyright law, which implicates a broader array of personal interests of the creator than patent law, may benefit from the application of the utilitarian framework to the extent that society seeks the production and diffusion of literary and artistic works. Hadfield (1992) provides a thorough historical survey of economic theories of copyright; see also Goldstein (1995, ch. 5) and Plant (1934b). The utilitarian framework has been particularly central to the development of copyright law in the United States. The Congressional Committee reporting on the 1909 Copyright Act stated: ‘The enactment of copyright legislation by Congress under the terms of the Constitution is not based upon any natural right that the author has in his writings, ... but upon the ground that the welfare of the public will be served ... by securing to authors for limited periods the exclusive rights to their

2. Historical Background

Utilitarian theories of intellectual property developed and evolved in a symbiotic relationship with the evolution of the modern state: from the formation and maturation of the mercantilist nation-states through the Industrial Revolution to the rise of the modern capitalist economy. Most early scholars focused upon what Merges (1995b) calls the ‘Grand Question’: whether state-created intellectual property rights should exist at all. More recently, attention has shifted toward the design of intellectual property rules and institutions.

Intellectual property rights emerged during the early mercantilist period as a means for nation-states to unify and increase their power and wealth through the development of manufactures and the establishment of foreign trading monopolies. The term patent, derived from the Latin *patere* (to be open), refers to an open letter of privilege from the government to practice an art (MMLJ, 1997, p. 122). The Venetian Senate enacted the first patent statute in 1474 providing the maker of any ‘new and ingenious device ... reduced to perfection so that it can be used and operated’ an exclusive license of 10 years to practice the invention. Other nations followed suit and the granting of limited monopolies for inventions, and later to publishers and authors of literary works, became the dominant means of promoting innovation and literature (Hadfield, 1992; Merges, 1995b; MMLJ, 1997).

The philosophy of intellectual property developed in response to the use of monopoly power to spur innovation. Adam Smith (1776, pp. 277-278), while generally critical of monopoly power as detrimental to the operation of the ‘invisible hand’, nonetheless justified the need for limited monopolies to promote innovation and commerce requiring substantial up-front investments and risk. Jeremy Bentham (1839, p. 71) went beyond this justification for intellectual property rights, providing a clear explication of the differential fixed costs borne by innovators and imitators:

> [T]hat which one man has invented, all the world can imitate. Without the assistance of the laws, the inventor would almost always be driven out of the market by his rival, who finding himself, without any expense, in possession of a discovery which has cost the inventor much time and expense, would be able to deprive him of all his deserved advantages, by selling at a lower price.
John Stuart Mill (1862) concurred that patent monopolies were justified, arguing that a temporary ‘exclusive privilege’ was preferable to general governmental awards on the ground that it avoided ‘discretion’ and ensured that the reward to the inventor was proportional to the ‘usefulness’ to consumers of the invention.

Pigou (1924) elaborated the basic framework of modern welfare economics, developing the concept of public goods as ‘instances in which marginal private net product falls short of marginal net social product because incidental services are performed to third parties from whom it is technically difficult to exact payment’. This appropriability problem figured directly in Pigou’s assessment of intellectual property:

The patent laws aim, in effect, at bringing marginal private net product and marginal social net product more closely together. By offering the prospect of reward for certain types of invention they do not, indeed, appreciably stimulate inventive activity, which is for the most part, spontaneous, but they do direct it into channels of general usefulness. (p. 151)

Clark (1927) reinforced this justification, noting that a system that did not give inventors control of their inventions would result in a rivalry in waiting for others rather than an effort to distance others in originating improvements.

Building upon the growing understanding of oligopoly and the economics of imperfect competition (see, for example, Robinson, 1933; Plant, 1934a, 1934b) offered a more skeptical view of intellectual property rights, questioning whether such rights were in fact needed to stimulate inventive activity and investment in actual as opposed to idealized markets. Plant argued that much invention is spontaneous and hence forthcoming without the provision of patent protection. He contended further that first-mover advantages, imperfections in markets and other factors provided inventors and publishers sufficient rewards to create and market their works even in the absence of intellectual property rights. Plant concluded that patent protection would lead to an overinvestment in research and development that could result in discoveries that fell within the patent domain, wastefully diverting resources from more appropriate endeavors.

Arrow (1962) provides the seminal modern diagnosis of markets for information. In addition to the appropriability problem described by Pigou, Arrow recognized that the marginal cost of increasing the utilization of information is zero.

Any information obtained, say a new method of production, should, from the welfare point of view, be available free of charge (apart from the costs of
transmitting information). This insures optimal utilization of the information but of course provides no incentive for investment in research. In a free enterprise economy, inventive activity is supported by using the invention to create property rights; precisely to the extent that it is successful, there is an underutilization of the information. (pp. 616-617)

Reflecting the Chicago tradition of law and economics, a number of scholars questioned whether the ‘public goods’ attribute of information is the most appropriate starting point for thinking about intellectual property. Demsetz (1969, 1970), applying insights from the property rights literature (Coase, 1960; Demsetz, 1967), argued that strong property rights for intellectual creations should be provided, with the market available to ensure efficient allocation of resources through Coasean bargaining. Hirshleifer (1971) undercut the public goods justification for intellectual property protection directly by pointing out that innovators may be able to appropriate substantial return from the private utilization of proprietary information without the need for property rights by speculating in markets on the basis of their discoveries prior to such discoveries becoming public knowledge. This mechanism creates strong incentives for the dissemination of such information following the speculative investing of the innovator.

By the late 1960s, economists increasingly turned their attention to the more narrowly focused question of how intellectual property rights should be designed to best promote innovation. In what is now considered the classic treatment, Nordhaus (1969) showed formally how the optimal duration of patent protection balanced the incentives for innovation against the deadweight loss of monopoly exploitation. Among his findings were that the optimal patent life is longer the lower the price elasticity of demand for the underlying product, the smaller the social benefit from the invention relative to the research and development cost, and the more responsive the amount of invention to the research and development cost. See generally Scherer (1972); Scherer and Ross (1990, p. 625).

3. Current Research on the Economics of Innovation and Intellectual Property Protection

By the early 1970s, three distinct models of intellectual property had developed: (1) a market failure framework building upon traditional neoclassical analysis; (2) a property rights framework reflecting the Chicago tradition; and (3) a comparative institutional perspective premised upon a
contextual analysis of the opportunities to appropriate the value of innovation in actual markets.

These models of intellectual property were static and focused upon the setting of initial property entitlements. A large body of institutional and historical work, however, emphasized the dynamism of innovation. A newer body of literature has introduced dynamic aspects of the innovative process directly into models of intellectual property. In addition, the modern literature has, as a result of the particular features of telecommunications, computer, and internet technologies, gone beyond the public goods rationale to examine the implications of network externalities for intellectual property protection. Moreover, scholars have recognized the role of market and social institutions in augmenting and tailoring intellectual property rules (for example, through contracting, joint ventures, and hybrid licensing institutions). They have also looked at other incentive structures - such as prizes, subsidies, and regulation - that provide alternatives and complements to intellectual property rights. Some of this work develops comparative institutional frameworks in order to assess the choice among institutions.

3.1 Social Value of Innovation

Robert Solow (1957) demonstrated that technological advancement and increased human capital of the labor force accounted for most (between 80 and 90 percent) of the annual productivity increase in the US economy between 1909 and 1949, with increases in the capital/labor ratio accounting for the remainder. Denison (1985) extends and refines this analysis, reaching similar results for the period 1929-1982: 68 percent of productivity gain due to advances in scientific and technological knowledge, 34 percent due to improved worker education, 22 percent due to greater realization of scale economies, and 13 percent attributable to increased capital intensity; these factors were offset by decreases in work hours (425 percent), government regulation (44 percent), and other influences. It is now widely recognized that technological advancement and enhanced human capital are the principal engines of economic growth in the United States and other industrialized countries (Scherer and Ross, 1990, pp. 613-614).

3.2 Historical, Industry and Institutional Studies

Joseph Schumpeter was among the first twentieth-century economists to recognize the fundamental importance of technological change in modern capitalist economies (Schumpeter, 1934, 1942, 1950; Nelson and Winter, 1982, pp. 275-281; Nelson, 1997). Schumpeter’s work emphasized three principles: (1) innovations continually upset established relationships in markets and organizational structures through a process of ‘creative destruction’; (2) technological innovation provides the opportunity for
temporary monopoly profit, and this linkage explains the rapid economic growth of the Western economies; and (3) large monopolistic firms are the prime source of technological innovation because they are best able to bear the high costs of technological innovation (Merges, 1988, p. 843). Although economists who study innovation generally accept Schumpeter's first two principles, most empirical studies of the relationship between market structure and research and development expenditures reject the linkage between monopoly power and disproportionately large investments in innovation (Scherer and Ross, 1990, pp. 614-660; Kamien and Schwartz, 1982, pp. 49-104). According to Scherer and Ross (1990, p. 660),

[S]chumpeter was right in asserting that perfect competition has no title to being established as the model of dynamic efficiency. But his less cautious followers were wrong when they implied that powerful monopolies and tightly knit cartels had any stronger claim to that title. What is needed for rapid technical progress is a subtle blend of competition and monopoly, with more emphasis in general on the former than the latter, and with the role of monopolistic elements diminishing when rich technological opportunities exist.

Economic historians have sought through case studies of particular innovations and industries to understand better the linkage between research and development and social welfare. These studies of the innovation process find that inventions are highly interdependent: 'Technologies ... undergo ... a gradual, evolutionary development which is intimately bound up with the course of their diffusion' (David, 1985, p. 20). Secondary inventions - including essential design improvements, refinements, and adaptations to a variety of uses - are often as crucial to the generation of social benefits as the initial discovery. See, for example, Nelson and Winter (1982), Taylor and Silbertson (1973), Mak and Walton (1972), Rosenberg (1972) and Enos (1958). Many studies emphasize the critical importance of linking innovation with understanding of consumer needs and astute marketing. Rosenberg (1976), for example, notes the dual emphasis that Thomas Edison placed upon technical research and understanding the existing market for illumination in bringing electrical lighting to market. Edison 'deliberately patterned many of his practices upon those of the gas industry. Edison’s commercial genius resides in an extremely shrewd awareness of those respects in which innovation called for continuity as well as discontinuity' (Rosenberg, 1976, p. 75; see also Freeman, 1980, p. 124, finding innovators’ attention to the education of users, publicity, market forecasting and understanding of user requirements to be key factors in distinguishing between successful and unsuccessful innovations). Saxenian (1994) describes how the process of diffusion depends not only upon downstream
considerations but also the degree of integration with component suppliers, the sharing of information and know-how within and among companies, industrial structure and a broad range of other upstream factors. These studies emphasize the importance of diffusion in the realization of social benefits from innovation.

Economic historians and industrial organization economists conducted a series of empirical studies during the 1970s and 1980s which attempted in part to assess the importance of intellectual property rights in spurring technological advance (Mansfield, 1986; Schwartzman, 1976; Taylor and Silbertson, 1973; compare von Hippel, 1988, pp. 46-53, reviewing other studies). These studies found that patents were rarely the principal means of appropriating returns in most industries (outside of pharmaceuticals and chemicals). In many industries, first-mover advantages, including the establishment of production and distribution facilities, rapid progress down a learning curve and other factors have proven to be at least as important as formal patent rights. Through a large survey of research and development personnel across a wide range of industries in the United States, Levin et al. (1987) found that intellectual property rights, in comparison with trade secrecy, lead time, rapid movement down the learning curve and marketing efforts, play a relatively modest role in enabling most firms (with the exception of those in the pharmaceutical and chemical industries) to appropriate returns for their inventions. Similar results have been found in Japan and Germany (Japan Institute of Intellectual Property, 1994; Oppenlander, 1984). Taken together, these studies suggest a growing consensus among economists that intellectual property rights offer a real, but limited, incentive to innovate in some industrial sectors, the importance of such rights vary significantly across industries and fields of innovation and the linkage between intellectual property rights and social welfare improvement is extraordinarily complex (David, 1985, 1993; Machlup, 1968; Merges, 1995b, pp. 107-108; Scherer, 1980; Sirilli, 1987; Stoneman, 1987, p. 115; Teece, 1986).

3.3 Enrichment, Refinement and Extension of the Economic Models
The early formal models of intellectual property rights assumed that inventors conducted research in isolation on non-competing projects. As such, the models focused on the optimal reward to induce particular innovations by a single inventor. As the historical and institutional literatures reflect, however, the actual environment for innovation is substantially more complex. In particular, different innovators (and firms) often compete to invent first, thus resulting in patent races. Economists needed to develop richer, dynamic models in order to understand the positive and normative implications of rivalrous competition. In addition, the traditional model of protection focused narrowly upon one instrument - patent duration - as the primary means for optimizing policy. Subsequent
studies have broadened the range of attributes of patent rights that can be varied, including patent breadth and exclusivity. Furthermore, most of the traditional models assumed that innovation resulted in an end product or process that could not be improved upon, that is, one-shot inventions. As the historical literature highlights, reflecting Sir Isaac Newton’s modest aphorism ‘If I have seen further than other men, it is by standing on the shoulders of giants’, most inventions are not only outputs but also inputs to the creative process, with subsequent innovators building upon a growing foundation. The modern literature has developed new models to study the implications of cumulative innovation for the design of intellectual property systems. The modern literature has also incorporated the insights of research on network externalities into the analysis of intellectual property protection.

Rivalrous Competition Economists have sought to understand whether the dynamics of rivalrous competition exacerbate or ameliorate the public goods problem associated with innovation. Positing a model in which the cost of innovation declines over time, Barzel (1968) showed that competition among innovators for a patent will tend toward too rapid innovation, while monopolist innovators who do not face potential entrants will tend to innovate too slowly. Barzel’s work inspired an extensive literature on rivalrous competition, which is ably summarized in Kamien and Schwartz (1982); Reinganum (1983) and Tirole (1988).

Gilbert and Newberry (1982) extend Barzel’s basic model by showing that monopolists may have an incentive to maintain their monopoly power by patenting new technologies before potential competitors can gain a foothold in the market (see also Cave, 1985; Gilbert and Newberry, 1984; Salant, 1984. As Tirole (1988) notes, two factors operate: an efficiency effect and a replacement effect. Because monopoly profits will be higher in any given market than oligopoly profits, a monopolist’s incentive to remain a monopolist (and hence pursue research and development intended to maintain such position) tends to be greater than a potential entrant’s incentive to become a duopolist (the efficiency effect). On the other hand, a monopolist has less to gain from innovation since it already is earning high profits. Innovation by the monopolist may only displace all or part of the existing monopoly profits (the replacement effect). Hence, monopolists may not have as much to gain as potential entrants from innovating. In the case of drastic innovation, the replacement effect dominates because the entrant becomes a monopolist (Reinganum, 1983). In the case of non-drastic innovation, the efficiency effect tends to dominate and hence monopoly power tends to persist (Fudenberg and Tirole, 1986).
The rivalrous competition literature, while quite diverse, has produced two general views of patent races: one suggesting that they inefficiently duplicate costs and the other suggesting that they encourage higher aggregate investment (thereby offsetting to some extent the public goods problem associated with information markets: Scotchmer, 1996b, 1998; Dasgupta and Maskin, 1987). The inefficiency branch emphasizes the duplication of research expenditures resulting from competition and decentralized research programs (Tandon, 1983; Wright, 1985). Rivalrous innovation produces an externality: by increasing the probability of invention (and obtaining a patent) through research and development efforts, a firm is thereby reducing the probability that its competitors will succeed in making the same invention (Loury, 1979). This effect can result in overinvestment in research and development. The efficiency branch emphasizes the way in which patent races accelerate the rate of investment and create pressure for continuing advancement (Fudenberg et al., 1983; Grossman and Shapiro, 1987; Harris and Vickers, 1987).

Extension of the Nordhaus Model The classic Nordhaus model assumed that only one policy instrument was available to encourage innovation through the provision of intellectual property rights: the duration of protection. Subsequent research has extended Nordhaus’s model, showing that the optimal duration of patents is longer where enforcement is costly or incomplete (Scherer, 1984) and where compulsory licensing at low rates is possible (Tandon, 1982). Optimal duration of patents is shorter where rivalry among firms raises the cost of innovation (McFetridge and Rafiquzzaman, 1986) and where prospective competitors may waste resources inventing around patents (Gallini, 1992 see also Chou and Shy, 1993).

Broader Range of Policy Instruments More recent work has focused upon other policy instruments, including the standard for protection, the scope of protection, the extent to which firms may collaborate (for example, joint ventures, licensing) and the enforcement of intellectual property rights (that is, property rules versus liability rules - compulsory licensing). Many of these studies have also incorporated the insights of the rivalrous competition literature and investigated the interaction of policy instruments. Due to the wide range of factors and interactive effects, this literature has raised at least as many questions as it has answered (see Priest, 1986). Nonetheless, this research has refined understanding of key interactions and the complexity of intellectual property policy.

Using a dynamic model of innovation in which firms decide what to patent and whether to engage in a patent race, Scotchmer and Green (1990) find that a weak novelty standard generally dominates a strong standard by
encouraging firms to disclose their discoveries. In addition, their model shows that a first-to-file rule for determining priority results in earlier disclosure of innovation than a first-to-invent rule, but it also creates excessive incentives for firms to pursue a patent race. O’Donoghue (1996), Scotchmer (1996a) and Luski and Wettstein (1995) extend the analysis of the standard for protection.

Numerous scholars have examined the implications of varying the scope (or breadth) of intellectual property protection. Gilbert and Shapiro (1990) show that the optimal breadth of patents should be extremely narrow and the optimal length infinite where wider breadth increases deadweight loss due to consumers substituting out of the product class. Klemperer (1990) finds that broad patents of short duration dominate long-lived, narrow patents where substitution to alternative products (within the same product class) is the main source of deadweight loss. Where potential competitors have a choice between waiting for a patent to expire and inventing around the patent, Gallini (1992) demonstrates that optimal patent scope should be broader (and duration shorter) in order to discourage prospective competitors from engaging in wasteful efforts to invent around patented inventions. Lerner (1994a) presents empirical evidence showing the importance of patent scope.

Collaboration and sharing of innovation, either through ex ante cooperative research arrangements or through ex post licensing agreements, provide a direct means to internalize the spillover benefits of research and development. Katz (1986), Grossman and Shapiro (1987), Scotchmer and Green (1990), Ordover (1991), Jorde and Teece (1992) and Gandel and Scotchmer (1993) discuss how collaborative research ventures address the appropriability problem and promote diffusion of innovation. They also note the anticompetitive risks of such ventures and propose antitrust tests for balancing competing considerations, emphasizing that such collaboration should be particularly encouraged with regards to basic research for which the appropriability problem is most pronounced (see also Kaplow, 1984). A number of studies examine the strategic licensing of innovation and how the potential for licensing may affect the level of research and development investment (see generally Tirole, 1988, pp. 410-414). Katz and Shapiro (1985a) find that firms will be less inclined to license major innovations (that is, innovations affording an effective monopoly) than minor innovations where the innovator and the potential licensees are comparably efficient because of the potential for the innovator to derive monopoly profits. They also find that the possibility of licensing has an ambiguous effect on research incentives. Whereas the potential for greater return (through more efficient diffusion) made possible by licensing encourages research, the returns to not innovating are also higher since losers of patent races (and those who do not enter the race) have the potential to share in the rewards so long as they have some bargaining power in negotiating a
license. Gallini and Winter (1985) describe two distinct motivations for licensing: (1) to avoid redundant costs; and (2) to discourage a competitor from pursuing research. These effects imply that licensing encourages research when firms' initial costs are relatively symmetric and discourages research when costs are asymmetric. Rockett (1990) explains how patentees may strategically license their innovations to ‘weaker’ competitors so as to prolong their dominant position in an industry even after a patent has expired.

As a means of addressing the anticompetitive effects of intellectual property protection, various economists and legal scholars have examined the use of alternative remedies for enforcing intellectual property rights. A number of studies have suggested that compulsory licensing of patents might be appropriate in particular circumstances - such as abuse of monopoly power (Adelman, 1977; Scherer, 1977, 1980; Tandon, 1982; Kaplow, 1984; Chang, 1995), and where network externalities exist (Menell, 1987) - but enthusiasm for this policy approach has been dampened in the United States by the determined opposition of industrial groups and the patent bar (Scherer, 1980, p. 456).

Scotchmer (1996a, 1998) notes that the effectiveness of the patent system could be improved by considering research and development costs of particular innovations in determining validity, duration and breadth of intellectual property protection, but recognizes that administrative and institutional constraints - such as the costs of verifying firms’ data and the accounting problem of allocating costs among products, projects and overhead - effectively preclude patent authorities from considering such data in practice.

*Cumulative Innovation* The classic Nordhaus model and much of the subsequent literature has assumed that research and development yield a single stand-alone invention which cannot be improved upon (and does not contribute to other research endeavors). As historical and industry studies have emphasized, however, relatively few innovations are pioneering. In addition, most inventions build upon existing work. Moreover, pioneering inventions often spawn entire new industries, with many improvements upon and new applications of the technology. Recent research has incorporated the cumulative nature of innovation into models of intellectual property protection.

Scotchmer (1991, 1996a, 1996b, 1998) describes a fundamental tradeoff arising as a result of cumulative innovation. In order to reward first generation innovators sufficiently for inventions that may produce positive spillovers by enabling second generation inventions (improvements, new applications and accessories), first generation innovators should be able to
appropriate the value of second generation innovations. On the other hand, providing even a share of the returns to second generation innovation to the first generation innovator reduces the incentive for second generation innovators to pursue their research. This tension is abated to the extent that first generation innovators are best positioned to pursue second generation innovation or where collaboration (for example, joint ventures) brings first and second generation innovation within the same profit center. The cumulative nature of innovation unquestionably strengthens the case for allowing joint ventures, especially with respect to complementary products (Scotchmer, 1996a; Katz and Ordover, 1989). In practice, however, it is rare that one entity is best positioned to pursue all second generation projects. Furthermore, second generation innovators are not known (and cannot be knowable) before first generation research investments must be made. Yet, once first generation research investments are made, they are sunk costs which become irrelevant for bargaining over the division of profits from multi-generation innovation. Green and Scotchmer (1995) and Scotchmer (1996b) suggest that this problem can be addressed by expanding the duration and scope of first generation patents or by denying patent protection altogether to second generation innovation. These results, however, depend critically upon strong assumptions relating to licensing of innovation and the knowledge and rationality of innovators. As the licensing literature notes, there are many strategic impediments to licensing of innovation. In addition, much of the institutional literature casts doubt on the degree to which innovators possess good information for assessing the best diffusion path for their technologies and whether innovators behave rationally in licensing to actual and potential competitors (Lemley, 1997b).

A number of scholars have extended the analysis of cumulative innovation. Chang (1995) shows that patent scope should afford greater breadth to inventions that have little value standing alone relative to the value of improvements. Without such protection, there will be insufficient incentive for innovators to pursue projects that will not yield substantial returns until second and later generations of improvements and/or applications have been developed. Matutes, Regibeau and Rockett (1996) show that patent breadth, rather than duration, is the key variable for promoting prompt disclosure of innovations. O’Donoghue, Scotchmer and Thisse (1997) show that providing broad patent scope improves diffusion of new products whereas patents of longer duration (and narrower breadth) tend to reduce R&D costs (see also O’Donoghue, 1996).

**Network Externalities** In addition to the public goods problem, many markets in which innovation plays an important role feature positive network externalities. A network externality arises where one consumer’s
use of a product increases the value derived by other users of the product. For example, adding an additional person to a common telephone network not only enables the new user to contact others already on the network, but also increases by one the number of people whom existing users of the network may contact. Similar positive externalities arise indirectly, for example, where software developers produce more application programs for widely distributed operating systems. As a result of advances in telecommunications, computer and internet technologies, network externalities are an increasingly important feature of modern economies.

Katz and Shapiro (1985a) have demonstrated that a new entrant to a market might adopt a noncompatible product standard even though their adoption of a compatible standard would increase social welfare. This behavior is driven by possible strategic advantages of not enhancing the desirability of the rivals’ products to consumers valuing standardization. Farrell and Saloner (1985) have shown a countervailing dynamic whereby the developers of improved standards may be unable to attract consumers because of the high switching costs to shift to the new standard. A large literature has developed examining the network externality phenomenon (see, for example, Farrell and Saloner, 1986, 1988, 1992; Gandel, 1994, 1995; Greenstein, 1993; Katz and Shapiro, 1986, 1992, 1994; Klemperer, 1987a, 1987b; Liebowitz and Margolis, 1995a, 1995b).

Drawing upon this literature, Menell (1987, 1989, 1998) shows that intellectual property protection has important implications for the dynamics of network externalities by affecting the extent to which competitors can establish proprietary standards. In markets featuring strong network externalities, Menell (1987) argues that the threshold for intellectual property protection should be higher than in traditional market settings so as to foster the adoption of standardized interfaces. In addition, compulsory licensing may be justified in particular circumstances to enable the full realization of network externalities. Intellectual property protection may also be particularly important in network markets so as to provide adequate rewards for firms to pursue research and development of improved standards, that is, to overcome bandwagon effects discouraging the development of improved standards (see also Farrell, 1989; Dam, 1995; Lemley, 1996; Lemley and McGowan, 1998).

3.4 Alternative Instruments and Comparative Institutional Analysis
A broad range of theoretical and empirical studies have looked at alternatives to traditional intellectual property protection as a means of encouraging innovation. These studies examine alternatives in isolation and comparatively (for example, Wright, 1983). All of these means have advantages and drawbacks in particular contexts. As with comparative institutional analysis more generally, no instrument or combination can
achieve perfection (Dwyer and Menell, 1997; Komesar, 1994). The goal is to identify that configuration of instruments and institutions which is most effective and responsive to the evolving nature of the problem. In addition, scholars have begun to study broader issues relating to innovation such as the role of regional economic structures and the way in which institutions evolve to address market failures in the intellectual property field.

As noted above, joint ventures and licensing arrangements provide an important means of sharing the costs of innovation and reducing the economic inefficiency resulting from spillovers and patent races, although they create conditions that may facilitate anticompetitive collusion among competitors and other problems. Prizes and tournaments may also provide strong incentives for research, but they also result in duplicative work (Carlton and Perloff, 1991; Rosen, Nalebuff and Stiglitz, 1983; Wright, 1983).

Institutional economists have focused more generally upon corporate strategies for appropriating returns from research and development. Business economists have identified a broad array of managerial choices affecting the rate and commercial success of innovation within a firm: trade secrecy, intra-firm competition, the use of stock options and other incentive-based compensation systems, suggestion boxes, marketing and licensing, strategic partnering, among other techniques for appropriating a return to investment. Using case studies and surveys, a growing literature in the business strategy field assesses how these various options may be integrated most effectively (see Teece, 1986; Levin et al., 1987; Storper, 1996; Storper and Harrison, 1991).

Government subsidies can directly promote the development of particular technologies. Governments fund significant research and development through basic science grants (for example, medical research), subsidies to higher education and other research institutions, procurement (especially military, space and environmental), technology development programs (for example, shale oil, nuclear power), tax credits (for example, solar energy) and other funding programs. Such policies can misallocate resources, however, because the government lacks adequate information to allocate, manage and monitor the use of subsidies effectively. Moreover, the provision of subsidies by the government generates rent-seeking by potential recipients which wastes resources directly and distorts the allocation of the subsidies.

Governments also seek to spur innovation indirectly, especially in the environmental and occupational health and safety fields, through regulatory programs (Ashford and Heaton, 1983; Ashford et al., 1985; Banks and Heaton, 1995; McGarity, 1994). Such programs often require industry to meet particular regulatory requirements and develop and install improved technologies (for example, lower polluting automobiles). Such ‘command and control’ and ‘technology-forcing’ regulations have not proven to be particularly effective because of the information costs and limitations of the
regulatory authorities and the political economy (rent-seeking) of regulatory programs (Menell and Stewart, 1994; Sunstein, 1990; Stewart, 1981; LaPierre, 1977). Incentive-based regulatory programs - such as effluent taxes and marketable permit systems - are often more effective because they use market forces to allocate resources to address pollution problems, but they are limited by the availability of accurate and cost-effective monitoring technology and the political impediments to the adoption of such policies (Menell and Stewart, 1994).

An alternative line of inquiry explores the ways in which intellectual property rights may undermine progress in science. Sociologists have emphasized the role of social norms among scientists, especially in the pursuit of basic research, in motivating discoveries. Merton (1973) identifies four interrelated norms of the scientific research community - universalism (emphasizing the objectivity of science), communism (viewing discoveries as the result of collaboration and hence should be dedicated to the scientific community), disinterestedness and organized skepticism - and non-economic reward structures - publication, reputation, professional advancement, esteem - which promote research and prompt disclosure. The biomedical community, in particular, has developed strong norms promoting the sharing of research to promote progress and serve humanity. Eisenberg (1987, 1989) notes that these norms and structures may conflict with the requirements, motivations and institutional structures of the intellectual property system. Trade secret, for example, directly undermines the disclosure of research. Even the patent system may delay disclosure until a patent application can be made. In addition, the exclusivity of the intellectual property system discourages sharing of discoveries and thereby slows the process of cumulative innovation by limiting access to scientific discoveries (Merges, 1996b). At a more general level, the increasing emphasis upon profiting from more commericalizable research in the universities has in part supplanted the more traditional drive to make pathbreaking basic discoveries. Others scholars emphasize, however, that intellectual property rewards are needed to diffuse academic and medical discoveries beyond the research community (Rosenberg and Nelson, 1994). Litman (1990) highlights the social and economic importance of a rich public domain to the creation of literary and artistic works.

Heller (1997) suggests that economic efficiency may be undermined by excessive division of property rights and the resulting bargaining breakdowns that can result from such fragmentation. Eisenberg and Heller (1997) examine the problems that are beginning to emerge in biomedical research as a result of a proliferation of intellectual property rights relating to the human genome. Lemley (1997a) addresses the increasingly prevalent problem of overlapping copyrights on the Internet.
In one of the most penetrating modern analyses of the determinants of innovation, Saxenian (1994) provides a detailed, multidisciplinary account of the factors driving the remarkable success of Silicon Valley as a robust and resilient engine of technological progress. Unlike most economic models of industrial organization economists which assume the basic preconditions of innovation (for example, rationality, information, contracting, competition), Saxenian pays close attention to the actual institutional forces motivating and sustaining rapid technological progress. She finds that Silicon Valley has thrived as a result of an extraordinary confluence of factors: a unique culture of collaboration and sharing of know-how both within and across firms; a high mobility of labor; competitive rivalry among many dynamic competitors; low barriers to entry; a high density of complementary specialized enterprises; the development of effective trade associations and consortia; creative use of strategic partnering, cross-licensing, second-sourcing and joint ventures; a responsive, knowledgeable and competitive venture-capital financing network that is integrated into the technology community; close research university-industry relationships; a legal culture emphasizing informal, practical, flexible and less litigious solutions (see Suchman and Cahill, 1996); a management style emphasizing teamwork, openness, participation and autonomy of decentralized engineering teams; and the use of stock options to attract and motivate employees and reward innovation. These factors have fostered sustained rapid technological progress and relatively stable economic growth in Silicon Valley, defying the predictions of product cycle theory (positing that regions follow a pattern of innovation, growth, maturation and scale production and ultimate decline as production shifts to other, lower cost regions). See also R. Nelson (1993) and Lindvall (1992).

Institutional Innovation Another promising line of research has examined the process by which new institutions form to address the limitations and gaps of existing intellectual property regimes. Besen and Kirby (1989) and Besen, Kirby and Salop (1992) examine the development of copyright collective organizations and the manner in which they reduce transaction costs in the licensing of intellectual property. Merges (1996a) surveys the broad array of institutions that have developed to address the appropriability problems and transaction costs plaguing intellectual property markets. Applying the insights of the new institutionalism literature (Cooter, 1994; Ellickson, 1991; Greif, 1989, 1993.; North, 1990; Ostrom, 1990; Powell and DiMaggio, 1991; Williamson, 1985) to case studies of performing rights societies, patent pools and the Hollywood Script Registry, Merges shows that strong property rules promote the innovation of hybrid institutions that can most efficiently address the transaction cost and valuation problems inherent in the reallocation of intellectual property.
rights. He argues that the advantages of institutional innovation militate against legislatively or judicially determined ‘liability’ regimes such as compulsory licensing and provide a justification for more permissive antitrust treatment of collective rights organizations in many circumstances.

4. Normative Analysis of Specific Modes of Intellectual Property Protection

4.1 Patent

Standard accounts of the patent system have emphasized several features of the law that promote economic efficiency: legal protection for invention encourages investment; disclosure requirements enhance technological knowledge and spur further research; incentives to develop and commercialize research rapidly diffuse advancements (Machlup, 1958, 1968; Penrose, 1951). These accounts emphasize a reward theory, seeing the appropriability of economic returns from investment as the driving force behind technological innovation (Oddi, 1996, pp. 275-277; Grady and Alexander, 1992, pp. 310-313).

More recently, a number of scholars have developed more sophisticated theories of how the patent system can best promote social welfare. See generally Kitch (1998), Oddi (1996) and Dam (1994). Reflecting concern about the social costs of monopoly power, Scherer (1980, pp. 443-450) has refined the standard reward theory to emphasize that the patent system should focus upon rewarding only those inventions that would not be forthcoming (or would be substantially delayed) without patent protection. In this view, patent protection would only be available for those inventions that are induced by the patent system itself (see also Oddi, 1996, pp. 275-281).

Building on Barzel’s (1968) model of rivalrous competition, Kitch (1977) argues that the patent system in essence provides the first to claim an invention with a ‘prospect’ or ‘development rights’, authorizing the patent holder to ‘mine’ the claim exclusively. Kitch defends this function of the patent system as promoting efficient (non-rivalrous) exploitation of innovation opportunities: technological advancement will proceed more desirably in such a regime because the first prospector will have ‘breathing room’ to develop the claim without fear that rivals will preempt or steal the claim and the inventor will be able to coordinate the development process. The opportunity to license the technology enables the inventor to contract with entities that may be better able to develop the claim. The prospect theory thus turns importantly upon a smoothly functioning technology
licensing market and the capacity, foresight and rationality of prospectors to coordinate the development and diffusion of the technology.

Emphasizing the other branch of the rivalrous competition literature (which sees social benefits to rivalry) and building upon the work of institutional scholars finding that technological development was retarded in industries where broad patents were granted, Merges and Nelson (1990) argue that the patent system should foster a ‘race to invent’. Drawing on a wide range of empirical evidence and theoretical models of bounded rationality, they see vibrant competition as a more positive force in spurring invention, innovation and diffusion of technology than coordinated development by a single prospector. Because of significant transaction costs in technology licensing markets, cognitive limitations of innovators and the unknown nature of technological ‘prospects’, Merges and Nelson argue that the patent system should promote competition in second generation inventions and improvements by limiting the breadth of patents (compare Lemley, 1997b). One of the benefits of rivalrous competition is that multiple inventors working on the same problem often produce different valuable inventions.

More recently, Grady and Alexander (1992) have offered a ‘rent dissipation’ model of the patent system. Like Kitch, Grady and Alexander focus upon the social loss from duplicative investment in innovation. In their terms, the patent system should (and does) strive to maximize the benefits that society derives from innovation less the development costs. The patent monopoly represents a rent that competing inventors will pursue until much of the net social benefit is dissipated through duplicative investment. Such dissipation occurs at three stages: in the race to obtain the initial patent; in the competition to develop those improvements ‘signalled’ by the patent; and in investments in secrecy to prevent rivals from obtaining spillovers from research efforts. In this model, the case for patentability declines as the value of the invention increases so as to reduce the rent dissipation that might occur at the pre-invention stage. Grady and Alexander would provide broad patent scope to those inventions that ‘signal’ a large potential for improvement. Their implications, therefore, directly contradict the ‘race to invent’ model. Oddi (1996), Martin (1992) and Merges (1992) question the normative and positive basis of this theory.

A large literature has also developed applying utilitarian theories of intellectual property to the analysis of particular doctrines of patent law. Kitch (1966) analyzes the non-obviousness standard. Merges (1988) uses empirical and theoretical research on the economics of innovation to critique an emerging trend in the case law using secondary factors such as commercial success in assessing whether an invention constitutes non-obvious innovation. Eisenberg (1989) offers a refined analysis of how the
experimental use exception of patent law could better promote progress of
science without undermining primary incentives. Merges (1994a) examines
the bargaining dynamics created by the exclusivity of patent rights. Dam
(1994) and Schlicher (1995) apply economic analysis to the broad spectrum
of patent doctrines.

4.2 Copyright
As with patent law, standard accounts of copyright law see the provision of
exclusive rights to reproduce original expression as an essential means to
promote literary and artistic creativity and the diffusion of works. Beginning
with Plant (1934b), however, a number of scholars have questioned the need
for copyright protection on utilitarian grounds. Plant argued that being first
in the market, the desire of authors to have their works and ideas widely
distributed and other factors, provide adequate rewards for the production of
literary works without the need for copyright protection. Hurt and
Schuchman (1966) and Breyer (1970) have since refined, elaborated and
narrowed this argument, developing a rudimentary form of comparative
institutional analysis. Tyerman (1971) presents a careful critique of this
argument to which Breyer (1972) replies.

More recently, economists have developed formal models of the
economics of copying in order to assess the effects of copyright protection.
Johnson (1985) and Novos and Waldman (1984) have shown that
limitations on copying enhance consumer welfare and address the
appropriability problem under plausible market conditions. Liebowitz (1985)
points out, however, that these studies do not take account of publishers’
ability to appropriate returns indirectly through discriminatory pricing of
originals, such as charging higher fees to libraries and other institutional
purchasers (see also Ordover and Willig, 1978). This effect reduces,
although does not eliminate, the appropriability problem for some classes of
works such as journals. The advent of very low cost copying and distribution
technologies such as the Internet, however, reduces the efficacy of this
appropriability means.

At a more doctrinal level, legal and economic scholars have applied the
insights of economic theory to the analysis of copyright doctrine. Gordon
(1982) explains how the fair use doctrine serves as a effective means for
permitting uncompensated use of copyrighted material where the
transactions costs of licensing or other means of exchange would prevent a
transfer through the market (see also Fisher, 1988). Kreiss (1995) analyzes
the extent to which copyright doctrines foster the accessibility of works,
which bears upon the promotion of learning and knowledge. He argues for a
liberal construction of the fair use doctrine and caution in awarding
injunctive relief. Merges (1992) assesses the fair use doctrine as applied to
parodies within an economic framework. Dreyfuss (1987) examines the work for hire doctrine of copyright law. Menell (1987, 1989, 1998) analyzes the efficacy and limitations of copyright protection in addressing the market failure associated with the creation and diffusion of computer software. Lunney (1996), elaborating Plant’s (1934b) principal observation, argues for a narrow scope of protection for copyrighted works (limited to exact or near-exact duplication) on the ground that broader protection would inefficiently promote investment in copyrightable endeavors at the expense of other activities in the economy. Lemley (1997b) suggests that copyright doctrines may discourage the creation of improved works. Economic analysis has also been used to analyze the efficiency with which copyright law achieves non-utilitarian goals, such as redistribution and the protection of moral rights. Numerous studies have analyzed the *droit de suite*, a continuing property right artists maintain in works of art which has been adopted in a minority of nations (Karp and Perloff, 1993; Hansmann and Santilli, 1997; Perloff, 1998).

### 4.3 Trademark

A trademark is a legally protectable name, word, symbol, design, or combination which designates the manufacturer of a product or service. The primary justifications for trademark law are ‘to facilitate and enhance consumer decisions’ and ‘to create incentives for firms to produce products of desirable qualities even when these are not observable before purchase’ (Economides, 1988, p. 526, 1998; McClure, 1979, 1996). A principal benefit of trademark protection is to lower consumer search costs (Carter, 1990; Landes and Posner, 1987; McCarthy, 1944, §2.1). Trademark law encourages manufacturers to invest in the development of brand names and distinctive packaging by eliminating the risk that competitors will free-ride upon such investments. The availability of reliable signals of product source and quality fosters informed consumer decisionmaking.

Some early industrial organization economists were critical of advertising (and hence marking) on the ground that they ‘unnaturally’ stimulated demand, thereby fostering and perpetuating oligopoly through ‘artificial’ product differentiation (Robinson, 1933, p. 89; Rosen, 1978; McClure, 1979, 1996; summarizing arguments). This view has been largely supplanted by theoretical arguments and empirical evidence supporting the view that advertising and trademarks are an efficient means of providing information in the marketplace (Stigler, 1961; Hirshleifer, 1973; Nelson, 1974; Nagle, 1981; Milgrom and Roberts, 1986; Landes and Posner, 1987; P’ng and Reitman, 1995; McClure, 1996).

Landes and Posner (1987) describe the manner in which various doctrines of trademark law promote economic efficiency (see also Folsom and Teply, 1980; Swann, 1980; Mims, 1984; Burgender, 1985). Much of
Landes and Posner’s analysis assumes that the set of marks available to a market entrant is practically infinite and hence there is no loss to the public when a mark is taken. Carter (1990) questions this assumption, noting that particular or a limited range of words or symbols are cheaper or more effective at branding some types of products and services than others. He analyzes the implications of a less than perfectly elastic supply of marks for some markets.

Brown (1948) notes that trademarks (and advertising more generally) create value by affecting the satisfaction consumers derive from acquisition and consumption of the good. Commentators have increasingly recognized such other effects of trademarks. Dreyfuss (1990) and Kozinski (1993) note that many trademarks have developed a stand-alone value - consumers value displaying a particular trademark, such as a team or corporate logo, wholly apart from any product or service that might be manufactured by the trademark owner. Kozinski argues on utilitarian grounds that the law should protect some prestige marks - such as Rolex - to foster the image advertising that generates such value. In order to maximize this value, some have argued that courts should protect an inherently distinctive trademark or trade dress from its inception in order to encourage the creation of valuable marks. Some also argue that trademarks should be recognized more directly as property rights (and not merely as a form of tort law) and be protectable against dilution (and not merely trading by competitors) (Schecter, 1927; Cullmann, 1947; Pattishall, 1976, 1984; Kitch, 1990; Swann and Davis, 1994). Carter (1990, 1993) and Port (1993, 1994) criticize these emerging trends in the law. Port, in particular, argues that trademarks should not be seen as property, but rather grounded in and limited to a narrow body of tort law.

4.4 Trade Secret
Innovators may seek to protect their intellectual work through the law of trade secrets, which draws from common law property, contract and tort doctrines and state statutes (see generally Merges et al., 1997, ch. 2). Trade secret protection directly addresses the appropriability problem by limiting contracting parties’ use and dissemination of proprietary information, thereby enhancing incentives to produce valuable information (Friedman, Landes and Posner, 1991). It is particularly important in encouraging the production of information that is not patentable, too expensive to patent, or more valuable if kept secret than protected through the patent system (Friedman, 1998). Trade secret law also serves to reduce the cost of keeping information proprietary by affording formal legal protection to those who make reasonable efforts to maintain trade secrecy. The subject matter of trade secrets includes almost any valuable information that is not generally known or available, for which the right holder is protected against misappropriation (acquisition by improper means or unauthorized disclosure) so long as he or she makes reasonable efforts to maintain secrecy.
of the information. Hence trade secrets can entail significant costs to maintain secrecy and negotiate, specify and monitor contractual relationships (Cheung, 1982). In addition, unlike patent law, the holder of a trade secret is not protected against independent discovery or reverse engineering. Moreover, the rightholder may have little effective recourse if a trade secret becomes widely known or if a user of the information acquired it without knowledge of its having been misappropriated.

Focusing upon the broader question of the comparative desirability of trade secrets within the broader regime of intellectual property protection, Cheung (1982) notes that trade secret protection inhibits the spread and use of ideas and causes the dissipation of economic rents. Nonetheless, trade secrets are an essential element of any economic system built upon freedom of contract and private property and they may be particularly effective (and more so than patent law) in promoting continuing research by employees or teams. Friedman, Landes and Posner (1991) suggest that trade secret law may be less prone to spurring over-rapid innovation (that is, excessive efforts to make a discovery, especially where the cost of an invention falls rapidly over time) than patent law because the negative consequences of coming in second place in such a ‘race’ are smaller. In a patent race, the ‘winner’ obtains an exclusive right to practice the invention, whereas in a trade secret regime, all independent discoverers are free to practice an invention.

Much of the economic literature on trade secrets addresses the optimal level of expenditures to maintain secrecy, that is, the question of reasonable efforts. Kitch (1980) argues that all such ‘fencing costs’ are inefficient and would require only such expenses as are necessary to provide evidence of the existence of a trade secret, that is, a notice or marking function (see also Reichman, 1994). Friedman, Landes and Posner (1991) make the related point that trade secret protection should be available when it is cheaper than the physical precautions that would be necessary to protect a particular piece of information. Another commentator argues that the optimal level of precaution against disclosure is at the point where the marginal cost of such measures equals the marginal expected loss in the event of misappropriation (Note, 1992). Friedman, Landes and Posner (1991) also defend the permissibility of reverse engineering under trade secret law as efficient. Scheppele (1988), however, argues that cases involving trade secrets are better explained in terms of fairness principles than in the efficiency terms of law and economics.

Lerner (1994a) presents empirical evidence finding trade secrecy to be a particularly important form of intellectual property protection for many businesses and one of the most frequently litigated intellectual property claims (see also Levin et al., 1987). Saxenian (1994) finds that a permissive
attitude toward the enforcement of potential trade secret rights has encouraged rapid diffusion and meteoric economic growth in Silicon Valley.

4.5 Misappropriation, Idea Protection, Right of Publicity and Sui Generis Forms of Protection

In addition to federal intellectual property protections (patent, copyright and trademark) that have developed in most industrialized and emerging economies, a range of state protections for intellectual work has evolved as well. As is discussed below, these protections are limited to some extent by federal supremacy and preemption. Nonetheless, they remain an important part of the broader intellectual property regime and have been responsive to new appropriability problems and gaps of federal systems. Such protections emanate from state legislative enactments as well as the development of common law doctrines in the courts. In addition, federal governments have augmented the troika of intellectual property law with more specialized forms of protection.

Baird (1983) and Raskind (1991) describe how state misappropriation doctrine, a form of tort law, has flourished and can be seen as reflecting competitive norms (see also Reichman, 1994). Lichtman (1997) argues on the basis of economic criteria that states should be able to craft limited forms of idea protection regimes that complement the federal regime. He argues that such protections will not undermine federal patent protection so long as such statutes limit innovators’ reward to their development costs.

Over the past two decades, state courts have developed and expanded the right of publicity, which protects celebrities from the having their image, voice, or other distinctive characteristic used by others. Grady (1994) argues that the right of publicity can be understood as a response to the problem of rent dissipation: free use of a celebrity’s attributes would lead to overuse, reducing the value to society. Using a congestion externality rationale, Posner (1992, p.43) argues that the right of publicity should be perpetual in duration: '[W]hatever information value a celebrity’s endorsement to consumers would be lost if every advertiser can use the celebrity’s name and picture'. Madow (1993) and Kozinski (1993, p. 975) argue on the basis of economic incentive and other considerations that the right of publicity has been expanded too far.

Merges (1996a) discusses the broad range of private intellectual property institutions that have evolved, including copyright collectives such as the American Society of Composers, Authors and Publishers (ASCAP) and Broadcast Music Incorporated (BMI) which license most musical works (see also Besen and Kirby, 1989; Besen, Kirby and Salop, 1992; Kobayashi and Yu, 1995; Koboyashi, 1998, patent pools and the Hollywood Script Registry.

4.6 Channeling Among Modes of Protection
Given the broad array of modes of intellectual property protection, each with differing standards and terms of protection, the overall efficacy of the intellectual protection regime depends significantly upon the ability of the system to properly channel innovation among the various modes. Four sets of doctrines - election, functionality, preemption and misuse - have developed to ensure that the overall system functions coherently.

The disclosure requirement of patent law in essence requires innovators to elect between patent protection and trade secret protection. As Goldstein (1974) points out, however, such election is only after-the-fact since inventors cannot perfectly predict whether their research will result in patentable inventions at the outset of their research efforts. Friedman, Landes and Posner (1991) argue that inventors use trade secret protection where innovations are either too trivial to be patented or where the costs of patenting outweigh the benefits.

Functionality doctrines serve to ensure that the exacting standards of patent law are not undermined by the bestowing of effective protection for the functional features of a work through other, less exacting, forms of intellectual property protection. The idea-expression dichotomy excludes functional features from protection under copyright law (Landes and Posner, 1989). Goldstein (1989, §2.3.1.1) suggests that courts engage in a rough balancing of the dangers of overprotecting and underprotecting a particular
work. Karjala and Menell (1995) explain how this principle efficiently channels protection between copyright and patent law generally and with regard to computer software. The functionality doctrine of trademark law limits the scope of trademark and trade dress protection to non-functional attributes so as to prevent a manufacturer from obtaining protection for functional features without meeting the requirements and being subject to the limitations of the patent law (Davis, 1996; Landes and Posner, 1987; Mims, 1984).

Preemption doctrines limit the authority of state legislatures and courts to develop protections within the fields occupied by federal intellectual property protection. Such doctrines not only override conflicting bodies of law, but also, through a negative implication, preempt the enactment or development of state doctrines to protect works that were left unprotected by federal law. Within the United States, federal patent, copyright and trademark law have broad preemptive domain, substantially limiting the authority of state legislatures and common law courts to develop local intellectual property regimes (Heald, 1991). Many scholars have criticized the breadth of such preemption, noting the lack of any empirical basis for preemption cases, the courts have allowed the development of a number of state intellectual property protections, including various misappropriation doctrines (including trade secret law), the right of publicity and limited forms of idea protection (Baird, 1983; Friedman, Landes and Posner, 1991; Heald, 1991; Raskind, 1991; Reichman, 1994; Lichtman, 1997).

Preemption doctrines also govern the relationship between state contract law and federal intellectual property law. Rice (1992) and Lemley (1995a, 1995b) argue that manufacturers of goods embodying intellectual property should not be permitted to extend the duration or scope of their rights through licensing provisions, particularly those specified in ‘shrinkwrap’ contracts. O’Rourke (1995, 1997) argues that copyright law should not preempt the enforcement of contractual terms that may alter parties’ rights with regard to copyrighted content (see also Hardy, 1995).

A fourth set of doctrines balances the policies of the intellectual property system, which grant limited monopolies in order to promote innovation, with the policies of the antitrust laws, which promote competition through the restriction of monopoly power. The intellectual property system would provide little spur to innovation if intellectual property owners were not permitted to exercise some market power. While the intellectual property laws create an implied limited exception to the antitrust laws, unrestrained freedom to exploit such monopolies could impose substantial indirect costs on society (over and above the deadweight loss attributable to monopoly
pricing), including the possibility that a patent holder could discourage further research in the field covered by the patent or seek to cartelize an industry through licensing agreements that foster collusion. A number of doctrines, including the patent misuse doctrine, have developed to address the difficulties at the intellectual property-antitrust intersection. The harmonization of the policies of these two bodies of law, however, is theoretically and practically quite complex. See generally Baxter (1966), Bowman (1973), Kaplow (1984), Teece (1986) and Lemley (1990) on patent misuse; Hanna (1994) on copyright misuse; Merges (1996a) on patent pools, and Merges et al. (1997, ch. 8).

5. Positive Analysis of Intellectual Property Protection

There has emerged in the literature two branches of positive analysis relating to intellectual property. As in many areas of private law, law and economics scholars reflecting the Chicago tradition have argued that intellectual property doctrine can be explained as a means for promoting efficient resource allocation. Kitch (1966, 1977) and Grady and Alexander (1992) have argued that patent law as applied by the courts has evolved toward an efficient set of doctrines relating to the standards for and scope of protection. Oddi (1996), Merges (1992) and Martin (1992) point out numerous inconsistencies with these claims. Landes and Posner (1987, 1989) have argued that the main contours of trademark and copyright promote economic efficiency. Grady (1994) makes similar claims for the right of publicity.

A second tradition, building upon the insights of public choice theory, has examined the political process producing intellectual property legislation and the extent to which such legislation reflects the outcome of interest group politics. In the view of one observer intimately familiar with the legislative process (Olson, 1989, p. 111),

Congress is generally not in the business of satisfying abstract concerns about ‘good copyright policy’. Rather, Congress is an intensely political body, loath to impose one-sided losses on legitimate interest groups. Since ‘good copyright policy’ would often require precisely such one-sided losses, copyright reforms may languish for decades before being enacted, or may simply be abandoned.

Litman (1989) describes the importance of interest group politics in affecting the way in which copyright law has evolved to accommodate and respond to technological change (see also Sterk, 1996, pp. 1244-1246). Menell (1994, pp. 2651-2652) highlights a perplexing dilemma in accommodating new technologies within the intellectual property system:
The opportunity for comprehensive reform is most propitious before interest groups form around a new technology. Unfortunately, policymakers usually do not have sufficient understanding of the path of such technology and the implications for an appropriate intellectual property regime during this nascent stage of development. Policymakers thus are left in the awkward position of either creating a regime before they adequately understand the problem or waiting until the contours of the problem emerge, at which point economic interests have vested and reform, if it is possible at all, is severely constrained.

Olson’s view that intellectual property legislation cannot move forward without consensus may be giving way to the view that those who stand to gain concentrated benefits from copyright legislation (content owners) may prevail in legislative fora (national and international) over those who stand to bear diffuse costs (consumers). Merges (1995b) notes a general shift in the societal baseline toward a presumption of protection (compare Lemley, 1997c; Karjala, 1987). He also observes that intellectual property rights are increasingly seen as an off-budget form of subsidies and hence they create strong incentives for interested parties to engage in rent-seeking. Karjala (1995), Lavigne (1996) and Sterk (1996) highlight the pressure to expand the term of copyright protection, notwithstanding any showing that such extension promotes creation of intellectual works. Commentators have criticized efforts by content owners to expand protections for their works on the Internet (Samuelson, 1996a). While the basic model of public choice is inadequate as a full explanation of the political process (see, for example, North, 1990, noting the vagaries of processes by which property rights are defined), it has important insights for understanding the evolution of intellectual property law.

B. Non-Utilitarian Theories of Intellectual Property

Non-utilitarian theories play an important role in justifying intellectual property. This is particularly true with regard to the protection of literary and artistic expression and publicity. The European nations have grounded intellectual property protection for such intellectual effort within non-utilitarian theories of rights. This difference in philosophical perspective is reflected in part in the ways in which intellectual property systems are designated. Whereas protection for literary and artistic expression in the United States comes within the ‘copyright’ law - the title of which emphasizes limits on the public’s right to make copies - the analogous bodies of law in Europe are labelled ‘author’s rights’: droit d’auteur in France, Urheberrecht in Germany and derecho de autor in Spain. Even in
the United States, however, there is a respected view to justify intellectual property law and to develop intellectual property doctrines on a broader base than the utilitarian model (Kaplan, 1967, p. 67; Treece, 1968; Hughes, 1988; Gordon, 1989, 1992a, 1992b, 1993; Yen, 1990; Netanel, 1993, 1996).

Over the past decade, a broad range of scholars has applied existing and novel philosophical frameworks to the analysis of intellectual property protection. Many of these scholars have drawn upon multiple philosophical strands in constructing their analyses. The discussion below disentangles the various strands so as to provide a comprehensive survey of the expanding foundation upon which intellectual property has been justified and critiqued.

6. Natural Rights/Labor Theory

John Locke offered a strong natural rights justification for private property which remains a central pillar of property theory today (Locke, 1698; Dwyer and Menell, 1997). Beginning with the proposition that all humans possess property in their own ‘person’, Locke argued that

\[
\text{[t]he ‘labour’ of his body and the ‘work’ of his hands, we may say, are properly his. Whatevsoever, then, he removes out of the stat that Nature hath provided and left it in, he hath mixed his labor with it and joined to it something that is his own and thereby makes it his property. It being by him removed from the common state Nature placed it in, it hath by this labour something annexed to it that excludes the common right of other men. For this ‘labour’ being the unquestionable property of the labourer, no man but he can have a right to what that is once joined to, at least where there is enough and as good left in common for others.}
\]

For elaboration of Locke’s general theory, see Simmons (1992) and Waldron (1979).

intellectual property rights should be absolute (see Waldron, 1993). Denicola (1981) and Ginsburg (1990) argue that copyright law should be interpreted broadly to allow protection for compilation of facts, even if they do entail original expression in their organization, so as to protect the ‘sweat of the brow’ inherent in creating such works (see also Ginsburg, 1992). Harris (1996) discusses the application of Lockean theory to ownership of body parts and products (for example, cell lines). Hettinger (1989) critiques the Lockean labor theory as applied to intellectual property, arguing that creators should be limited in their property interest to the value they add by applying their labor to things removed from the commons and not to the total value of the resulting product (see also Nozick, 1984, pp. 175-182).

7. Unjust Enrichment

Gordon (1992a, 1992b) argues that the central problem of intellectual property law is to compensate creators of works who bestow benefits on those who follow, up to some socially justifiable point. In this analysis, the basic structure of intellectual property law is closely akin to the law of restitution, which seeks to determine when someone who bestows unbargained-for benefits deserves compensation.

8. Personhood Theory

The personhood justification for property derives from Kant’s Philosophy of Law and Hegel’s Philosophy of Right and has been elaborated in modern legal discourse in the work of Radin (1982, 1993). The premise underlying the personhood perspective is that to achieve proper development - to be a person - an individual needs some control over resources in the external environment. The necessary assurances of control take the form of property rights’ (Radin, 1982). The personhood justification for property emphasizes the extent to which property is personal as opposed to fungible: the justification is strongest where an object or idea is closely intertwined with an individual’s personal identity and weakest where the ‘thing’ is valued by the individual at its market worth. For general critique of this theory, see Schnably (1993), Simmons (1992).

Netanel (1993) traces the rich heritage of Continental copyright law and its moral rights tradition to the personality theory developed by Kant and Hegel, pointing out nuances distinguishing the various strains within the theory (see also Palmer, 1990, pp. 835-849). For example, Kant viewed literary work as part of the author’s person and hence is not alienable.
Hegel, by contrast, distinguished between mental ability as an inalienable part of the self, but not the act of expression. Netanel presents a multifaceted argument for alienability restrictions upon copyright interests. The broader implications of the personhood justification for intellectual property have been explored by a number of scholars: Hughes (1988) (suggesting various strains of the personhood theory in American copyright law); Port (1994) (disputing Hughes’ use of personhood theory to support anti-dilution actions in trademark); Cherensky (1993) (with regard to works for hire); Hughes (1998) (right of publicity); Solomon (1987) (right of publicity). Personhood theory has been particularly central to the emerging debate, brought to the fore by advances in biotechnology, over property rights in body parts, cell lines and other body products (Munzer, 1994; Radin, 1987).

9. Libertarian Theories

Palmer (1989, 1990) constructs a libertarian argument against intellectual property rights by critiquing the dominant philosophical perspectives used to justify intellectual property protection. Coming from a different intellectual tradition, but reaching a similar conclusion, Barlow (1994) argues that intellectual property rights threaten to undermine free exchange of ideas over the Internet and enable corporate interests to exercise substantial control over cultural and political expression. Netanel (1996, pp. 365-385) suggests that these concerns can be addressed better through reworking rather than discarding copyright law. More generally, Waldron (1993) points out that autonomy as an ideal cuts both for and against intellectual property rights. Authors may claim that the integrity of their self-expression requires that they control the use and adaptation of their works. Social commentators may argue, however, that they are denied the ability to express themselves if they cannot parody the works of others.

The issue of personal autonomy also arises with regard to control of body parts and cell lines. Informed consent may address this concern, although the meaning of informed consent may be more subtle in the context of modern medical research, where scientists may be able to decode a patient’s genetic structure and produce valuable byproducts (Harris, 1996; Lavoie, 1989). In addition, as above, liberty interests do not decisively cut in just one direction.
10. Distributive Justice

Theories of distributive justice seek to distribute society’s resources on the basis of just principles. Many philosophers endorse utilitarian theories of distributive justice (Mill, 1862; Singer, 1975; Hare, 1978). Such theories often reflect Lockean and other philosophical perspectives as well (Sterk, 1996, pp. 1234-1239). The process of determining such principles is the focus of considerable debate among political philosophers. Rawls (1971, 1993), for example, offers an ‘ideal contractarian’ theory of distributive shares in which a just allocation of benefits and burdens of social life is determined by what rational persons would choose from behind a ‘veil of ignorance’, which prevents them from knowing what abilities, desires, parentage, or social stratum they would occupy. Firth (1952) rejects contractarian approaches and instead argues that justification derives from a suitably defined Ideal Observer. Nozick (1984) approaches such questions from a non-interventionist standpoint, arguing that the State should play no role in distributing or redistributing property come by properly apart from respecting the voluntary transfers of property owners.

Considerations of distributive justice have recently been applied in justifying intellectual property. Rakowski (1991, pp. 86-87), for example, develops a rich theory of justice with applications to the distribution of the rewards of invention. Landes (1992) invokes Rawls’s ‘veil of ignorance’ to argue that authors as a group would favor limited ‘productive uses’ of unpublished materials within the scope of fair use. Similarly, Brennan (1993) uses Rawls’s contractualism to suggest a just set of rules to govern the fair copying of expression in works of criticism. Sterk (1996, pp. 1234-1239) argues that copyright law lacks coherence by reference to the leading competing theories of distributive justice.

The most concrete manifestation of distributive justice principles in the intellectual property field are recent international accords with regard to the protection, ownership and use of resources. Advances in biotechnology have spurred the prospecting of biological resources throughout the world, which has increasingly brought the traditional principles and values of the intellectual property system (emphasizing scientific and technological advance through limited, exclusive monopolies) in conflict with larger social justice, sovereignty and access concerns (Sedjo, 1992; Kadidal, 1993; Goldman, 1994; Carroll, 1995; Urbanski, 1995; Adair, 1997). The International Undertaking on Plant Genetic Resources of the United Nations Food and Agriculture Organization provides specifically for ‘farmer’s rights’. It recognizes farmers as innovators entitled to intellectual integrity and access to the germplasm and technologies they have developed collectively over many generations. Such recognition serves to protect the
interests of those who may lack the knowledge or resources to perfect their intellectual property rights against the exercise of formal rights by better organized agricultural interests that develop patent portfolios. The Convention on Biological Diversity provides that parties cooperate to ensure that intellectual property rights ‘are supportive of and do not run counter to’ the objectives of the convention, the conservation and sustainable use of biodiversity and the equitable sharing of benefits. The Convention also requires signatory nations to ‘respect, preserve and maintain knowledge, innovations and practices of indigenous and local communities’.

In addition, biotechnology has also led to the development of valuable resources derived from the human body. Distributive justice arguments have been used in assessing moral claims to cell lines and other products of the human body (Harris, 1992, 1996; Rakowski, 1991, pp. 167-195).

11. Democratic Theories

Copyright law promotes political expression by encouraging expression, but it also potentially inhibits dissemination of works by prohibiting, subject to some limitations, the copying of expression. In its early history, copyright was used by the English Crown to regulate the press (and censor seditious expression) through bestowing selective royal grants of privilege (Goldstein, 1970; Kaplan, 1967). Although copyright no longer functions directly to censor political expression, it nonetheless has the potential to inhibit the free flow of information. Goldstein (1970) discusses how the principles of copyright law - including the idea-expression dichotomy, the fair use doctrine and a misuse doctrine - harmonize this body of law with constitutional protections of freedom of speech and the press (see also Nimmer, 1970; Denicola, 1979; Patterson, 1987). Coombe (1991) offers a post-modernist critique of intellectual property law, arguing that the expanding domain of intellectual property protection limits the ability of individuals to express themselves. Netanel (1996) suggests that copyright plays an increasingly important role in modern democratic societies because of the ease with which expression can be disseminated through the use of digital technology. He argues that existing theories of intellectual property rights may undermine larger democratic principles and articulates a new model for the interpretation of copyright in the digital age which seeks to promote a democratic civil society.
12. Radical/Socialist Theories

A radical critique of some basic assumptions underlying intellectual property - most notably, the romantic concepts of ‘the author’ and ‘the inventor’ - has developed in recent years, building upon the work of deconstructists in the field of literary criticism. These scholars suggest that the concept of authorship and inventorship is so malleable, contingent and ‘socially constructed’ that we should be wary about identifying a creative work too closely with a particular person or entity (Aoki, 1993-1994, 1996a, 1996b; Boyle, 1988, 1992, 1996; Jaszi, 1991; Lange, 1992; Woodmansee and Jaszi, 1994). According to this view, all creations are the product of communal forces to some extent. Dividing the stream of intellectual discourse into discrete units, each owned by and closely associated with a particular author or inventor, is therefore an incoherent exercise subject more to the political force of asserted authors’ or inventors’ groups than to recognition of inherent claims of natural right, personhood, or other justifications. Lemley (1997c) and Samuelson (1996a) assess and critique the main claims of the deconstructionist view.

13. Ecological Theories

As understanding of the connections between technology, industrial development and the environment has deepened, environmentalists have begun to view intellectual property within the context of broader philosophical theories relating humans to the environment. Traditional reformist theories - which are anthropocentric in nature and draw heavily upon utilitarian, liberal and other traditional philosophical frameworks - see technology as both a source of environmental problems and a means for reducing environmental impacts of human activities. These theories generally call for the internalization of the adverse impacts of technology and the use of subsidies and other mechanisms (such as intellectual property) to spur the development of new technologies that reduce environmental impacts (see generally Banks and Heaton, 1995; Menell and Stewart, 1994).

Environmentalists have developed naturalist ethics over the past 50 years which challenge key assumptions of traditional philosophical perspectives (Nash, 1989). Most significantly, these theories are built upon non-anthropocentric premises. Leopold (1949) developed an ecological ethic, which seeks to ‘preserve the integrity, stability and beauty of the biotic community’. Leopold argued that the complexity of the biosphere counsels caution in any activities that might disrupt natural processes (see generally Devall, 1980). Singer (1975) called for recognition of moral considerateness
of all sentient beings. A broad range of environmental activists, Green Party politicians and social commentators have used these perspectives to attack technological advance as a social goal generally (see, for example, Commoner, 1971). Of particular significance in recent years, many ecologists and animal rights activists have questioned the encouragement of biotechnology, both in terms of the risks of adverse environmental consequences from the release of novel genetic organisms and the morality of re-engineering living organisms.

C. Conclusions

14. Synthesis and Future Directions

Intellectual property is rarely justified on one theory, although patents’ grounding in utilitarianism comes the closest. Consensus about philosophical perspective, however, has not produced consensus about what that perspective prescribes. Economic theorists have produced multiple plausible models for which empirical distillation will remain elusive and unlikely to be of much general predictive value due to the heterogeneity of inventive activity, the diversity of research environments, the complexity of technological diffusion, the richness and changing nature of real world institutions and the obvious measurement problems in conducting empirical research of this type. The comparative advantages of various configurations of intellectual property rights, antitrust standards, government subsidies, regulation and other encouragements for innovation are difficult to assess. The operation of these various alternatives turn on key assumptions - such as the extent to which firms will aggressively innovate without direct competitive pressure and the impediments to licensing, joint ventures and other transactional mechanisms (ex ante and ex post) - for which empirical evidence is limited in general and with regard to heterogenous contexts in which these issues arise. Nonetheless, the work of the past decade has refined some of the debates significantly and provided valuable, although sobering, evidence on others. Of perhaps greatest importance, recent work has shown that the holy grail of a perfectly calibrated incentive system is unattainable. Especially when the insights of public choice theory are factored into the analysis - in particular, the difficulty of enacting and implementing public-regarding intellectual property policies in the presence of rent-seeking by interested parties - the field should focus more on setting the main parameters and providing incentives for the evolution of new privately and socially constructed institutions to develop effective governance structures.
When the theoretical domain is expanded beyond utilitarian analysis, as it is in some patent contexts and most other areas of intellectual property, scholars have looked principally for parallel implications and conflict among competing philosophical justifications as a means of assessing justifications for particular intellectual property rules and institutions. This pragmatic approach (Kaplan, 1967) rarely produces intellectual tidiness, but is an essential aspect of justifying governance regimes in diverse social, political and economic cultures. Many factors are at work, which leads to rules that channel protection among modes of protection and varies the thresholds for and nature of protection within particular modes. As technology advances, the system continues to evolve, sometimes by new legislation, more often by the stretching and bending of existing rules. New technology commercialized in the past two decades, most notably the advent and diffusion of digital technology and new advances in the life sciences, portend deepening interest in the intellectual property system and scrutiny, reconsideration and reconceptualization of the theories justifying intellectual property. Even within the existing theories of intellectual property, these technologies pose significant analytical challenges as a result of the ways in which they change key factors on which existing institutional rules and structures are based - for example, the nature of personal and liberty interests of creators and users, network dimensions, transaction costs. As intellectual property and technology have gained importance over the past two decades, the philosophical debates have melded with broader social and political discourse bearing upon the very foundation of modern society. We can expect that intellectual property will continue to press these frontiers as the information age progresses.

Acknowledgements

I am grateful to Nancy Gallini, Paul Goldstein, Mark Lemley, Rob Merges, Richard Nelson, Eric Rakowski, Suzanne Scotchmer, Ralph Winter and an anonymous referee for helpful discussions and comments on an earlier draft, although they bear no responsibility for any shortcomings.
Bibliography on Intellectual Property: General Theories (1600)


Carleton, Dennis W. and Perloff, Jeffrey M. (1991), Industrial Organization, Glenview, IL, Scott Foresman.


Economides, Nicholas (1988), ‘The Economics of Trademarks’, 78 Trademark Reporter, 523-539.


Freeman, Christopher (1980), The Economics of Industrial Innovation, Cambridge, MA, MIT Press.


Leopold, Aldo (1949), A Sand County Almanac, New York, NY, Oxford University Press.


Mackaay, Ejan (1985), ‘De Hersenschim als Rustig Bezit - Moet Alle Informatie Voorwerp van Eigendom zijn? (The Chimera as Peaceful Possession - Must All Information be the Object of a Property Right?)’, Computerrecht, 12-16.


Mill, John Stuart (1862), Principles of Political Economy (5th edn), New York, Appleton.


Riis, Thomas (1994), 'Ofphavsret og Kulturekonomi i EF (Copyright and The Economics of Culture in EC)', NIR.
Riis, Thomas (1996), Ophavsret og Retsekonomi (Copyright and Law and Economics), Gadjura, PhD thesis.
Robinson, Joan (1933), The Economics of Imperfect Competition, London, Macmillan.


Scheppele, Kim Lane (1988), Legal Secrets: Equality and Efficiency in the Common Law, Chicago, IL, University of Chicago Press


