Intellectual Property and the Firm

Dan L. Burk†

Since its articulation by Ronald Coase, the economic theory of the firm has generated an enormous body of literature in corporate law and related fields.[1] That literature has become increasingly dominated by property-based theories of the firm that consider the allocation and disposition of institutional assets. Given this influence and direction, it is curious that the theory of the firm has received relatively little attention in intellectual property scholarship. In a so-called information age, where the most important assets of firms increasingly are intangible assets, one might expect that property-based theories of the firm would be readily applied to intellectual property.

While the dearth of such analysis might suggest that theories of the firm have little to tell us regarding intellectual property, this supposition is not only counterintuitive, but belied by scholarship employing these theories in the intellectual property context, which is now beginning to emerge. Some work in this regard has already been done by Robert Merges, considering the law of ownership applicable to inventions created by employees.[2] Although Merges does not explicitly invoke the theory of the firm, his analysis has elements common to such theories. Merges has also explicitly applied the theory of the firm to consider the control of innovators over inputs into their production.[3]

More recently, David McGowan has considered theories of the firm in relation to the licensing of copyrighted "open-source" software. He suggests that the purportedly decentralized open-source programming community operates more as an economic firm than romanticized visions of this community might suggest.[4] McGowan's

---

† Oppenheimer, Wolff & Donnelly Professor of Law, University of Minnesota. The author thanks Brett McDonnell, David McGowan, D. Gordon Smith, and participants in the University of Chicago Intellectual Property Colloquium for helpful comments in the formulation of this Essay.

1 See generally Jason Scott Johnston, *The Influence of* The Nature of the Firm *on the Theory of Corporate Law*, 18 J Corp L 213 (1993).

2 Robert P. Merges, *The Law and Economics of Employee Inventions*, 13 Harv J L & Tech 1 (1999) (arguing that economic theory justifies the legal rule denying employees ownership of innovations developed while on the job).

3 Robert P. Merges, *Intellectual Property Rights, Input Markets, and the Value of Intangible Assets* (draft Feb 9, 1999), online at http://www.law.berkeley.edu/institutes/bclt/pubs/merges/iprights.pdf (visited Dec 16, 2003).

4 David McGowan, *Legal Implications of Open-Source Software*, 2001 U Ill L Rev 241,

work in this context implies, though does not explicitly state, that intellectual property licensing may be a coordinating factor in the production function of economic firms. Likewise, D. Gordon Smith has considered fiduciary duties in light of property-based theories of the firm. Smith does not explicitly consider intellectual property, but his analysis clearly maps onto the law of trade secrecy.[5]

This scholarship suggests that there are insights to be gained by considering intellectual property in light of theories of the firm. However, as indicated, much of the work to date only hints at the fruits that such an exercise might bear. Even where the conjunction of the two is explicit, the commentators tend to take the state of intellectual property doctrine as given. Both Merges and McGowan are interested in the theory of the firm primarily as a vehicle to evaluate the innovative practices of certain industries, rather than as a tool to assess the state of intellectual property law itself.

In this Essay I wish to employ the theory of the firm in this latter fashion. I offer a preliminary examination of the relationship between intellectual property and theories of the firm, considering whether intellectual property law has progressed in the way we might have predicted in light of the theory of the firm, and to the extent that it has not, asking whether the theory of the firm has any recommendation to make for its improvement. In particular, I examine whether existing intellectual property law provides for efficient allocation of intellectual property rights within firms in a manner that comports with property-based theories of the firm. I begin with a short overview of property-based theories of the firm, and then suggest how such theories might fit one conception of intellectual property law. Finally, I focus briefly on doctrines in several major areas of intellectual property that suggest themselves for consideration in light of theories of the firm.

I. THEORIES OF THE FIRM

Modern theories of the firm, particularly the property-based theories considered here, have evolved in order to explain and justify the presence of organizational hierarchies within free market systems. But, paradoxically, the economic theory of the firm has yet to include any commonly accepted definition of the organization contemplated, that is, the concept of the "firm."[6] Certainly the term as used in this context is not synonymous with its meaning in everyday conversation

---

263–65.

[5] See D. Gordon Smith, *The Critical Resource Theory of Fiduciary Duty*, 55 Vand L Rev 1399, 1444–47 (2002).

[6] See Harold Demsetz, *The Economics of the Business Firm: Seven Critical Commentaries* 6 (Cambridge 1995) (noting that the literature lacks "precise definitions of the firm").

or even in legal parlance; the economic firm is not necessarily a legally recognized organization, although some legally recognized firms are also economic firms.[7] The economic literature on the firm generally uses the term to denote an area of economic activity characterized by hierarchical organization and command production, rather than by market negotiation.[8] The production center thus denoted is generally conceived to encompass an entrepreneur controlling a variety of inputs, including employees, via a complex of relationships, usually contractual in nature.[9]

This conception of the firm, derived from the early work of Coase and of F.H. Knight, exists to explain the presence of hierarchical organizations within free markets.[10] Coase, in particular, developed his early theory of the firm to explain why firms should exist even though markets are considered an efficient mechanism to coordinate productive activity. As in much of Coase's other influential work, the answer depends largely upon an inefficiency of markets, transaction costs that attend negotiated bargains. Firms, Coase postulated, exist in order to lower such transaction costs. Under this model, the productive activities of the firm are directed by an entrepreneurial fiat, rather than being realized via negotiated market transactions. In some cases, hierarchical production may prove less costly than market production due to the transaction costs of the market. In such cases, competitive pressures will tend to compel the formation of firms, as market participants organize themselves to minimize inefficiencies or face displacement by competitors that have already done so.[11]

This model predicts the emergence of firm organizations where the transaction costs of the market become too high. Commentators expanding upon Coase's insight have identified opportunism as one of the key transaction costs associated with bargaining in the marketplace[12] and have introduced theories of incomplete contracting into the analysis. Parties to a transaction cannot anticipate all future con-

---

7 See Eric W. Orts, *Shirking and Sharking: A Legal Theory of the Firm*, 16 Yale L & Pol Rev 265, 272–75, 312–14 (1998) (discussing distinctions between legal and economic firms).

8 See Ronald H. Coase, *The Nature of the Firm*, 4 Economica 386 (1937), reprinted in Ronald H. Coase, *The Firm, the Market, and the Law* 33 (Chicago 1988).

9 See Michael C. Jensen and William H. Meckling, *The Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure*, 3 J Fin Econ 305, 310–11 (1976).

10 See Coase, *The Firm, the Market, and the Law* (cited in note 8); F.H. Knight, *Risk, Uncertainty, and Profit* (Houghton Mifflin 1921).

11 See Oliver E. Williamson, *The Mechanisms of Governance* 233–34 (Oxford 1996).

12 See Oliver E. Williamson, *The Economic Institutions of Capitalism* 30 (Collier Macmillan 1985); Oliver Hart, *An Economist's Perspective on the Theory of the Firm*, 89 Colum L Rev 1757, 1762–63 (1989) (assessing Williamson's suggestion that independent contractors are more likely than employees to engage in hold-up behavior). As Harold Demsetz points out, Coase appears to regard opportunism as only one of the costs that might drive the formation of firms. See Demsetz, *Economics of the Business Firm* at 20 (cited in note 6).

tingencies and, as a consequence, any contract they negotiate will be incomplete.[13] Thus, contracts will necessarily present opportunities for one or the other party to take advantage of unforeseen developments. In particular, a party to a contract may attempt to "hold up" the other party, extorting additional concessions once resources have been committed to a project and cannot be easily recommitted to another venture.[14] Such relationship-specific resources, by virtue of their tailoring to a particular project, increase the efficiency of projects but also create this potential for opportunism.[15] Generalized resources may be more easily recommitted to new uses but are less well suited to any given project. Thus, in the face of possible hold-ups, parties may face an undesirable incentive to avoid asset specificity.

The development of a firm, which organizes production by command rather than by negotiation, may be explained as an attempt to deter such hold-up problems by eliminating repeated negotiations. If production is organized by executive fiat rather than by negotiation, there may be fewer opportunities for hold-ups to occur. However, by integrating production into a hierarchical structure, hold-up problems may simply be moved within the firm.[16] Employment relationships are themselves contractual, and the boundary between internal and external contracts may be difficult or impossible to define. Some commentators have suggested that the relationships that comprise the firm are subject to much the same incentives and characteristics as those between firms;[17] other commentators have extended this concept, modeling the firm as a nexus of contractual relationships, where team production is organized through a web of contractual arrangements.[18] At some point, these employment contracts spill over into arms-length relationships with independent contractors outside the firm. The resulting model portrays the firm in many ways as a microcosm of mar-

---

13 See Oliver D. Hart, *Incomplete Contracts and the Theory of the Firm*, in Oliver E. Williamson and Sidney G. Winter, eds, *The Nature of the Firm: Origins, Evolution, and Development* 138, 140–42 (Oxford 1993) (outlining the theory of incomplete contracts in the context of the theory of the firm).

14 See Lloyd Cohen, *Holdouts and Free Riders*, 20 J Legal Stud 351 (1991) (describing the "holdout" problem and distinguishing it from the "free rider" problem). See also Williamson, *The Economic Institutions of Capitalism* at 76, 133–35 (cited in note 12) (arguing that bringing transactions within a firm makes opportunistic behavior less likely).

15 See Benjamin Klein, Robert G. Crawford, and Armen A. Alchian, *Vertical Integration, Appropriable Rents, and the Competitive Contracting Process*, 21 J L & Econ 297, 298–300 (1978) (discussing asset specificity, opportunism, and the risk of asset appropriation).

16 See Hart, 89 Colum L Rev at 1763 (cited in note 12) (pointing out that scholars have yet to reveal the mechanism by which the incentive for opportunism is reduced by bringing services within a firm).

17 See Armen A. Alchian and Harold Demsetz, *Production, Information Costs, and Economic Organization*, 62 Am Econ Rev 777, 777 (1972).

18 See Jensen and Meckling, 3 J Fin Econ at 310–11 (cited in note 9).

ket relationships, reflecting in miniature the same contractual structures within and without the firm's boundary.

This suggests that the major transaction costs delineating the boundaries of the firm may be coordination and agency costs: the interests of employees may not be perfectly aligned with that of the entrepreneur directing firm activity, prompting opportunistic behavior. Employees may see opportunities to exploit situations not foreseen in their employment agreements, and, once assets have been devoted to a project, they may hold up the firm for additional concessions. Commentators analyzing this scenario have argued that opportunism can be deterred only by some form of "gap-filling" provision allocating control of resources in situations not contemplated by contract.[19] The analysis suggested by these commentators places property rights into this role, effectively to serve as the default for incomplete contracts.[20] Allocation of ownership rights to the firm is thus advanced to ameliorate the problem of opportunism within the firm; proprietary control of relationship-specific assets prevents employees from using them to hold up the firm.[21] Thus, under a property-based model, the right to exclude, and the concomitant right to grant access to the assets of the firm, ultimately define the boundary of the firm.

The types of assets contemplated by developers of this property-based model appear to have been the largely traditional physical assets of production: buildings, machinery, furnishings, paper clips.[22] Less tangible assets, such as specialized financial instruments, may also have been contemplated, and specific human capital has surely been a key consideration in the model's approach to hold-ups. But the assets of the modern firm are at least equally likely to accrue as intangible resources: ideas, know-how, information, inventions, goodwill, and the like. Thus, consonant with the transactional and property-based models reviewed here, the firm might be modeled as a nexus of intellectual property rights vested in entrepreneurs, together with actual or implied licenses based on those rights, permitting employees to use and modify the firm's intellectual property.

---

19 See Oliver Hart, *Firms, Contracts, and Financial Structure* 29–31 (Oxford 1995); Oliver Hart and John Moore, *Incomplete Contracts and Renegotiation*, 56 Econometrica 755, 755–57 (1988).

20 See Oliver Hart and John Moore, *Property Rights and the Nature of the Firm*, 98 J Polit Econ 1119, 1120–22 (1990).

21 See id at 1149–51. See also Raghuram G. Rajan and Luigi Zingales, *Power in a Theory of the Firm*, 113 Q J Econ 387, 423–24 (1998).

22 But see Hart, *Firms, Contracts, and Financial Structure* at 56 (cited in note 19) (mentioning intellectual property as a firm asset).

II. PROPRIETARY KNOWLEDGE

In the United States, intellectual property systems have most often been characterized as instruments to induce investment in creative products that, in the absence of a proprietary right, might be underproduced.[23] Creative products frequently display characteristics similar to those of "public goods," that is, goods that can be enjoyed by more than one individual simultaneously, and from which it is difficult to exclude consumers.[24] Since consumers cannot easily be excluded from enjoying the good, they would be unlikely to pay for it, and creators, knowing this, will be unlikely to invest in producing it in the first place. Proprietary rights give the creator a legal right to exclude, which allows the creator to derive an income stream from selling access to the work, either by selling the work directly or by collecting royalties from others who sell the work.

Typically, then, intellectual property rights have been viewed as the basis for market negotiation; by assigning exclusive interests in creative works, intellectual property facilitates bargaining.[25] In other words, intellectual property rights are conceived of primarily as mechanisms for coordinating activities *between* firms. However, as detailed above, the boundary of the firm is often indeterminate, and the distinction separating contractual relationships within the firm from contractual relationships between firms may be ambiguous. Property-based theories of the firm suggest that the right to access and use dedicated resources must be allocated within the firm as well as beyond the firm. This means that, in addition to their recognized inter-firm functions, proprietary rights may also serve to coordinate resources within a firm.

This function of proprietary rights in the firm has not been the focus of the standard explanations for intellectual property. However, at least one theory of intellectual property has focused on the assignment of proprietary rights as a mechanism to coordinate development of the creative product. This "prospect theory" of intellectual property is most closely associated with the work of Edmund Kitch, who argued that the assignment of intellectual property places development of valuable innovation in the hands of an entrepreneur who can then coordinate development of that resource.[26] Following this rationale, one may argue that by vesting firms with control of such intangible as-

---

23 See William M. Landes and Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J Legal Stud 325, 326 (1989).

24 See id.

25 See Frank H. Easterbrook, *Intellectual Property Is Still Property*, 13 Harv J L & Pub Pol 108 (1990).

26 Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J L & Econ 265, 276 (1977).

sets, the exclusive intellectual property rights found in patent, copyright, and trade secrecy may serve to prevent opportunism and promote coordination of intangible resources.[27] Employees or potential licensees who wish to develop proprietary information must seek the permission of the rights-holding entrepreneur. This view of the "prospect" coordination rationale is also consonant with models of the firm as a nexus of contractual production. Employees routinely make, use, sell, offer for sale, reproduce, distribute, adapt, and otherwise make use of the firm's intellectual property in ways that would constitute infringement if the uses were unauthorized. We typically presume that such uses are authorized but may have given little thought to the source of the authorization—the license that allows employees to make use of the firm's intangible assets in the course of their employment.

On this model, firms may be reluctant to develop project-specific intellectual property if control is incompletely allocated, as this sets the stage for potential hold-up by employees. Allocation of rights may be relatively obvious where tangible property or monetary assets are concerned; the employee who converts or appropriates such assets to his own use is guilty of theft, embezzlement, or waste. But misappropriation of intangible assets such as know-how, concepts, inventions, or designs is more problematic because the nonrivalrous nature of the resource leaves the appearance that nothing has been "taken." Indeed, if rights to intangible assets are left inchoate, employees may be unable to distinguish their personal knowledge and creative resources from those developed under the auspices of the firm.[28]

The point is perhaps most clear in the case of trade secrecy, which arises out of a complex pedigree of tort, contract, and equitable legal claims. Although trade secrecy disputes occasionally involve blatant industrial espionage, more typically they arise in cases of employee mobility, where a departing employee is purported to have taken proprietary information. Other typical cases involve instances of corporate divorce, where a former corporate ally is purported to have misused proprietary information after a joint venture or licensing relationship had gone sour. Resolving trade secrecy disputes is especially problematic in the case of employee departure, as courts are reluctant to curtail the mobility of labor. Society typically will benefit if workers are encouraged to acquire employable skills. But such general ability may be difficult to separate from the specific human capital in which

---

[27] See John F. Duffy, *Rethinking the Prospect Theory of Patents*, 71 U Chi L Rev (forthcoming 2004) (arguing that the "prospect" features of the patent system channel potential patentee rivalry, maximizing the social benefits from a patent monopoly).

[28] See Edmund W. Kitch, *The Law and Economics of Rights in Valuable Information*, 9 J Legal Stud 683, 709 (1980).

an employer has invested for a particular project. Thus both employee and venture situations are characterized by the need to identify and allocate assets that were specifically dedicated to the particular relationship.

As such, much of trade secrecy law is concerned with demarcation of firm assets, designating incentives and methods for giving proper notice to employees, partners, or licensees of the business information considered to be proprietary and confidential.[29] The substantive law reflects such attempts at demarcation. Trade secrecy is typified by a duty of confidentiality arising in a business relationship.[30] Confidence cannot be imposed upon another party without that party's implied or express consent, so the circumstances giving rise to trade secrecy must be such that a party knew or should have known of the confidentiality obligation.[31] Such obligations may arise from the formal statement of a written contract, or from a contract implied in fact based upon the behavior of the parties, or even from a contract implied in law based upon the general expectations of society.[32] Firms are also expected to take "reasonable measures" to ensure secrecy of their proprietary information[33]—for example, marking confidential information as confidential, password protecting or physically securing such information, and logging or monitoring use of the information.

Such measures signal to employees which assets are considered valuable and project-specific by the employer. Moreover, the additional investment in confidentiality measures is unlikely to be made in generalized assets that pose no danger of hold-up. Consonant with this principle, the law of trade secrecy excludes from protection information that is either generally known or readily ascertainable from information that is generally known. Neither does trade secrecy law protect against reverse engineering or independent discovery of the secret. Generally ascertainable knowledge, or knowledge that can be independently discovered at a low enough cost to make such re-creation feasible, is unlikely to be information that is project-specific, nor is such knowledge likely to represent enough of a project-specific investment to make hold-up likely.

Indeed, the remedies available for misappropriation of trade secrets might be characterized as designed to negate the benefits of hold-up, placing the opportunist monetarily or injunctively where he

---

29 See id.

30 See *E.I. du Pont de Nemours Powder Co v Masland*, 244 US 100, 102 (1917).

31 See Restatement (Third) of Unfair Competition § 41 (1993).

32 See id § 41(b)(1)–(2) (stating that a duty to maintain the confidentiality of trade secrets arises in situations where the circumstances would suggest to both parties that the disclosure was to be kept in confidence).

33 Uniform Trade Secrets Act § 1(4)(ii), 14 ULA 437, 438 (1990).

would have been without the benefit of the proprietary information. Courts may delay use of the misappropriated information for the period of time that it would have taken to reverse engineer or independently develop the secret, or deprive the wrongdoer of the profits gained by not having to expend resources on reverse engineering or independent development. Such head-start damages or head-start injunctions effectively place the opportunist where he would have been had he been operating wholly outside the boundaries of the firm. This characterization additionally implies that joint ventures or licenses that convey project-specific assets fall within the boundaries of the firm for purposes of the property-based theory, even though the entities involved may be separate actors for legal purposes.

Yet, paradoxically, trade secrecy does not confer a property right, or at best it confers an incomplete property right. Trade secrecy does not confer a right as against the world, but only as against a limited number of individuals that stand in a confidential relationship with the proprietor. Courts have displayed some confusion as to this quasi-property right, sometimes treating it as property and other times repudiating that label.[34] But under a firm-based analysis, this incomplete propertization is not a problem; confidential information does not confer a right as against the world, but it need not, and public welfare dictates that it probably should not. A full discussion of the economic and constitutional limits of trade secrecy lies beyond the scope of this Essay, but for this discussion, it is enough to recognize that trade secrecy need only confer an allocative property right against opportunism as between the parties to the secret.

III. COPYRIGHT

The discussion to this point suggests that many aspects of trade secrecy are consonant with the need to distinguish project-specific from general assets and to allocate specific assets so as to avoid opportunism—thus trade secrecy law aligns fairly well with the property-based theory of the firm. At the same time, trade secrecy is costly, requiring cumbersome secrecy measures and detailed confidentiality agreements—in other words, requiring parties to an employment contract to largely foresee the conditions of possible opportunism. Allocation of more complete rights via patent or copyright avoids the need for cumbersome secrecy or for ex ante contractual obligations to establish proprietary rights in information insofar as these systems en-

---

[34] See Pamela Samuelson, *Information as Property: Do* Ruckelshaus *and* Carpenter *Signal a Changing Direction in Intellectual Property Law?*, 38 Cath U L Rev 365, 365–67 (1989) (explaining the reluctance of some courts to characterize trade secrets as property and describing two recent Supreme Court decisions that do appear to classify information as property).

able the firm to establish an exclusive right in intangible relationship-specific assets.

In the case of copyright, this outcome has been statutorily formalized under United States law via the work made for hire doctrine. Under this doctrine, individual employees who create copyrightable works while operating within the scope of their employment are not considered to be the authors of those works. Rather, the institution employing the creator becomes the legally recognized author.[35] Thus, the key considerations determining whether a work is made for hire are the statutory criteria of "employee" and "scope of employment."

The Supreme Court has interpreted the statute to require that "employees" be distinguished from independent contractors on the basis of agency principles.[36] The factors used to assess an individual's employment status may include the ability of the worker to accept or decline assignments, the provision of materials by the employer, the duration of the relationship, tax treatment of the worker, and provision of benefits.[37] The test is flexible, with no particular factor designated as dispositive, although some lower courts have given greater weight to certain ubiquitous factors.[38] The scope of employment criterion is similarly judged under agency principles. Relying upon the Restatement of Agency, courts have defined the scope of employment in terms of the type of work the individual was hired to perform, the temporal and locational boundaries of the employment, and the motivation for creating the copyrightable work.[39] Under the third factor, the creative activity must have been motivated at least in part to serve the interests of the employer.[40]

The work is considered by statutory default to be made for hire whenever these criteria obtain. This has several legal consequences,[41]

---

35 17 USC § 201(b) (2000) (declaring that "an employer or other person for whom the work was prepared" is considered the author of a work made for hire). See also 17 USC § 101 (defining "work made for hire").

36 See *Community for Creative Non-Violence v Reid*, 490 US 730, 742–43 (1989).

37 See id at 751.

38 See *Aymes v Bonelli*, 980 F2d 857, 861 (2d Cir 1992) (holding that the *Reid* factors "should be weighed according to their significance in the case," and finding that five of the factors will be relevant in most cases); *Carter v Helmsley-Spear, Inc*, 71 F3d 77, 86 (2d Cir 1995) (noting that five of the *Reid* factors will be relevant in nearly all cases).

39 See *Avtec Systems, Inc v Peiffer*, 21 F3d 568, 571 (4th Cir 1994) (applying common law agency principles to determine if an innovation was created within the scope of a party's employment); *City of Newark v Beasley*, 883 F Supp 3, 7 (D NJ 1995) (applying a three-part test derived from the common law of agency to determine the scope of a party's employment).

40 See *Cramer v Crestar Financial Corp*, 38 USPQ2d 1684, 1689 (4th Cir 1995); *Food Lion, Inc v Capital Cities/ABC, Inc*, 946 F Supp 420, 422 (MD NC 1996), affd, 116 F3d 472 (4th Cir 1997).

41 For example, works made for hire are ineligible for certain moral rights protections under the Visual Artist Rights Act. 17 USC § 106A (granting the author of a work certain "moral" rights); 17 USC § 201(b) (designating the employer as the author in cases of work made for hire).

but for the present discussion, the most relevant consequence of work made for hire status is the default ownership of the work. Designating the employer as author means that ownership rests with the firm, subject to contractual transfer, rather than initial employee ownership subject to contractual transfer.

The copyright statute also allows designation of authorship for certain commissioned works.[42] In certain statutorily defined situations, the employee and employer may designate by contract whether the work will be considered a work made for hire. Thus, both ownership and authorship are open to negotiation in these situations. There is a good deal of historical path dependence to the statutory list of commissioned works. Industry organizations must have sufficient political capital to have a type of commissioned work added to the list, or, as in the case of sound recordings' brief appearance on the list, sufficient political capital to effect removal.[43]

But in general, the list contemplates subject matter that is highly collaborative or requires multi-party coordination: motion pictures, compilations, atlases. Such works are among the most likely to be subject to hold-up problems if property rights are fragmented. Such works are also most likely to involve negotiations on the border of the firm, where firms are engaged in recruiting contractors who may sometimes fall inside the firm's transactional boundaries and sometimes outside. In the latter case, market negotiations will be more efficient than entrepreneurial fiat. Some flexibility in adopting or eschewing the property default rule makes sense in these instances, as the parties will know best on which side of the firm boundary a contractor's activity falls.

Of course, copyrightable works created outside an agency relationship may always be assigned to an employer by means of an employment contract. This is the common practice in U.S. patent law, which lacks a work made for hire provision.[44] But reliance on contractual assignment re-introduces the problem of indefiniteness. Copyright, in particular, may foster ambiguities, as the statute requires a signed writing to transfer exclusive rights in a work,[45] purportedly in order to create certainty in transfers.[46] Blanket assignment of works in

---

42 17 USC § 101 (defining "work for hire" and allowing certain commissioned works to be designated as such if the parties agree in a written instrument).

43 See Robert A. Gorman and Jane C. Ginsburg, *Copyright: Cases and Materials* 282 (Foundation 6th ed 2002) (summarizing the brief history of Congress's addition of sound recordings to the statutory list of commissioned works and their subsequent removal).

44 See text accompanying notes 50–52.

45 See 17 USC § 204(a).

46 See *Effects Associates, Inc v Cohen*, 908 F2d 555, 557 (9th Cir 1990) (explaining that committing licenses or assignments to writing "prevents misunderstandings" and "enhances predictability and certainty of copyright ownership").

advance of an individual's general employment may not satisfy this statutory requirement, as the contract would not contemplate the transfer of any specific work. Pre-assignment might be more viable where an individual is engaged to create a specifically contemplated work, and this is indeed the situation under the "commissioned works" categories. One can imagine that a duty to transfer or a nonexclusive license might also be implied from the employment relationship, as has commonly been the case in patent law. Indeed, an implied nonexclusive license would look like patent law's shop right.[47] But all of these permutations create ambiguities that might prove opportune for hold-up.

Work made for hire sidesteps this possibility by creating an allocative default rule. In several dimensions, the conceptual structure of copyright's work made for hire doctrine appears a natural extension from the fiduciary-based principles of trade secrecy reviewed above. The doctrine has developed via application of agency principles, which, like fiduciary principles, have been identified as consonant with property-based theories of the firm. Tellingly, in the work made for hire situation, courts are not typically concerned with the classic agency situations—claims of third parties against a principal for actions of the purported agent or claims by principals against an agent for breach of a fiduciary duty. Rather, agency principles are used to allocate control of what Gordon Smith has called a "critical resource" in which the firm has invested, and in which we would wish to encourage specific investment.[48]

This division appears roughly consonant with the criterion of "asset specificity" under a property-based theory of the firm. The factors and categories defining employee works made for hire, or commissioned works, will tend to define situations where we would want to encourage development of project-specific resources. However, the doctrine as currently constituted may also allocate some nonspecific assets to the control of the firm. This occurs when authorship, and not simply ownership, is allocated to the firm. By erasing the identity of the natural creator, work made for hire removes from the natural author a reputational interest that is otherwise specific to the natural person, and not the firm. This interest is recognized most strongly in nations with strong moral rights regimes and less so in U.S. copyright. But work made for hire denies the natural author even the limited U.S. moral rights regime of the Visual Artists Rights Act.[49] Thus, an "asset specificity" approach suggests that authorship and ownership

---

47 See text accompanying notes 52–53.

48 See Smith, 55 Vand L Rev at 1444–46, 1455–57 (cited in note 5).

49 See note 41.

should perhaps be bifurcated under work made for hire, allocating the reputational interest to the natural author even while assigning default ownership of the work to the firm.

IV. PATENT LAW

Unlike copyright law, U.S. patent law lacks a work made for hire provision. Although corporations or institutions can be authors, they cannot be inventors; only natural persons can be so designated. This is something of a puzzle, as the need for a default rule of institutional ownership might seem more pronounced in patent law than in copyright. With a few exceptions, such as large motion picture projects, creativity within the traditional subject matter of copyright tends to be relatively low-cost; a modest investment in pencil and paper may suffice. Patent development may sometimes be similarly inexpensive. But corporate research and development is typically expensive, nearly always collaborative, and frequently critical to a firm's survival. These factors should militate toward developing the sort of clear ownership rule that has emerged in copyright, but oddly has not in patent.

Of course, one might argue that patent law has no need for institutional inventors because a patent law work made for hire doctrine is rendered unnecessary by liberal application of contract law. As a matter of employment practice, sophisticated research operations routinely require employees, as part of their employment agreements, to assign patents or other invention rights to the employer.[50] In theory, such "pre-assignment" contracts will create a duty for an employee to assign the rights to whatever may be discovered. But, as a practical matter, in a significant number of situations the contracts either are never executed or fail to anticipate the circumstances of invention, lending support to the concern regarding incomplete contracts. And, unlike copyright, the default rule for patent allocates property rights to the employee rather than to the firm.

In the absence of explicit contractual terms requiring an assignment, an implied duty to assign may be found. Courts have tended to recognize such an implied duty to assign patent rights in situations where an employee hired to solve a problem engages in research, and the invention relates to that effort.[51] Initially, courts recognized this duty only where employees were hired to solve a specific problem, but the rule has evolved to apply to employees engaged to perform gen-

---

[50] See Merges, 13 Harv J L & Tech at 7 (cited in note 2).

[51] See generally Steven Cherensky, Comment, *A Penny for Their Thoughts: Employee-Inventors, Preinvention Assignment Agreements, Property, and Personhood*, 81 Cal L Rev 597 (1993); William P. Hovel, Note, *Patent Ownership: An Employer's Rights to His Employee's Invention*, 58 Notre Dame L Rev 863 (1983).

eral research and development, as well as those engaged for specific research. In each case, invention was part of the employment duties undertaken, so a duty to assign the invention to the employer should come as no surprise to the employee.

In situations where an employee was not hired to engage in research, but developed an invention in which the employer has an interest, courts have assigned to the employer a nonexclusive license known as the "shop right."[52] The shop right, in some cases, is characterized as a contract implied in fact, inferred to exist in situations where the inventor appeared to acquiesce to the employer's use of the invention. In other cases, it appears as a license implied in law, arising from an equitable sense that the employer is entitled to use an invention developed on the employer's time by means of the employer's materials. This implied license evolved into an independent right of the employer, although a limited one: the right is personal and nonexclusive and does not pass by license or assignment.

Although the shop right arises in situations involving nonresearch employees, inventions subject to the shop right must still relate to the employment. In determining the existence of the shop right, courts have taken into account whether the invention relates to the duties of the employee, whether the invention falls within the scope of the employer's business, and whether the invention was created with materials supplied by the employer or was developed during working hours. When the invention falls within these criteria, the employer enjoys the shop right. The employee need not convey the full interest in the invention, however, due to what the courts perceive as the "peculiar nature of the act of invention,"[53] which courts characterize as springing from the employee's own inventive genius, and not from the material contributions of the employer.

This characterization of the inventive process appears out of step with the reality of industrial research and development, but it does suggest a possible reason for the absence of a work made for hire doctrine in patent law: a romantic notion of the inventor inventing in an individual flash of genius may be exerting an even greater influence upon patent doctrine than that of the romantic author upon copyright law.[54] The result of imposing the shop right is to effectively grant the employer a compulsory license at a zero royalty. The lack of exclusiv-

---

52 See Scott P. Sandrock, *The Evolution and Modern Application of the Shop Right Rule*, 38 Bus Law 953, 961–63 (1983) (outlining the evolution of the shop right rule in the context of employees who are not hired to perform research tasks).

53 *United States v Dubilier Condenser Corp*, 289 US 178, 188–89 (1933).

54 See Peter Jaszi, *Toward a Theory of Copyright: The Metamorphoses of "Authorship,"* 1991 Duke L J 455; Martha Woodmansee, *On the Author Effect: Recovering Collectivity*, 10 Cardozo Arts & Enter L J 279 (1992).

ity, however, leaves the employer at the mercy of the employee's residual control of the invention.

Professor Merges has suggested that these employee contract rules might be explained under contractual theories of "penalty defaults."[55] These theories share with property-based theories of the firm a pedigree rooted in concepts of incomplete contract. The general argument suggests that in the face of incompleteness, contractual defaults can play an information-forcing role.[56] Thus, the terms of the contract should be calibrated to penalize the party with the greatest informational advantage if that party is not forthcoming in the contracting process. In the specific case of employee inventions, the implication is that the employer, being the more sophisticated party to the contract, will be more forthcoming in negotiating invention assignments under the threat of a default rule that strips the employer of invention ownership.

But the information-forcing story is a largely unsatisfying explanation for the rules that have developed, supplying at best an incomplete justification of the employee ownership doctrine. There would seem to be relatively little information to force. The employment contract is signed at the beginning of the relationship between employer and employee, when the employer has relatively little information about the employee's abilities, creativity, or work habits. To be sure, the employer will have scrutinized the employee's references, resume, transcripts, and interview responses, but these are likely to provide little indication of the types of inventions the employee may develop, or the circumstances under which she may develop them. For that matter, invention is frequently serendipitous, defying premeditation and fostering the image of the romantic inventor. But we nonetheless would wish to encourage investment in the environment that fosters such serendipity, even if the employer has no advance information about the precise outcome.

The theory of the firm may supply a more complete explanation of employee ownership rules, setting default rules to ensure residual control rather than to force information disclosure. In developing employee ownership rules, courts have essentially dragooned doctrines of implied contract or fiduciary duty into bridging the gaps left by incomplete contracts. The doctrines play precisely the role contemplated by property rights under property-based theories of the firm. The criteria chosen for these employee rules suggest that courts are groping toward the demarcation line of asset specificity. Inventions created by

---

55 Merges, 13 Harv J L & Tech at 19 (cited in note 2).

56 See Ian Ayres and Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 Yale L J 87, 94 (1989).

employees specifically employed to do research are perhaps most likely to be those in which the firm has become irrevocably invested. Inventions created by non-research employees, even those using the firm's resources, may be those on which the firm is less likely to be dependent. As such, the latter inventions are less likely candidates for opportunism. As with copyright's work made for hire doctrine, an employee's activity within the "scope of employment" and use of firm resources are indicators of relationships lying within the boundaries of the firm, rather than out in the market.

At the same time, such judicial gap-filling is just that—attempts to fill gaps in incomplete contracts by inferring the existence of unarticulated agreements from the parties' actions, or creating obligations in law that appear to result in more-just outcomes. The solution itself is piecemeal, developed ad hoc, and riddled with discontinuity. While the case law tends toward stable designation of residual rights, it lacks the certainty of an actual property allocation.

This same uncertainty appears in the employment-related aspects of substantive patent law itself, especially in the concept of the "inventive entity."[57] Inventive entities may be composed of a single individual inventor or a group of inventors. Inventive groups composed of different individuals constitute different inventive entities even if their memberships partially overlap. In other words, an inventive entity composed of Alice, Bob, and Carol is not the same inventive entity as that composed of Alice and Carol, or of Ted, Carol, and Bob. Because patent law recognizes natural individuals, and not juridical individuals, as inventors, firms are not typically considered inventive entities. Instead, research groups within a firm are considered inventive entities.

Inventions must be novel and nonobvious to qualify for patent protection. The novelty requirement is typically operationalized under the statute by requiring that the invention cannot have been known or used by "others" in the United States.[58] But different inventive entities, even when including some of the same individuals, might constitute "others" for purposes of the statute. Thus, one research group could generate prior art that might interfere with the patentability of related inventions created by another research group composed of some of the same personnel. The inventive entity comprising Alice, Bob, and Carol might generate prior art against a related invention created by Bob, Carol, and Ted, or even by Bob and Carol. The common knowl-

---

[57] See generally Robert Patrick Merges and John Fitzgerald Duffy, *Patent Law & Policy: Cases and Materials* 766 (LexisNexis 3d ed 2002).

[58] See 35 USC § 102(a) (2000) (denying patents where the inventions have been previously known, used, or described in publication). See also 35 USC § 102(e) (declaring that novelty is negated if a patent application has been filed "by another"); 35 USC § 102(g)(2) (declaring that novelty is negated if there exists a prior invention "by another").

edge carried by natural persons into each of these inventive entities almost ensures overlapping discoveries, but any difference in membership causes that knowledge to be imputed to a different "inventor." Prior art generated by another overlapping entity could also be used to deny patents on grounds of obviousness, as prior art that qualifies for section 102 novelty purposes may also be used to measure obviousness.[59]

Courts recognizing this problem began by carving out exceptions where the outcome of this rule seemed unduly harsh.[60] Congress has addressed other parts of the problem through piecemeal legislation. For example, Congress has twice amended section 103 of the patent statute in order to exclude categories of subject matter from consideration as prior art if the subject matter and claimed invention were either commonly owned or subject to a common assignment.[61] The statutory amendments track the general rule against an inventor defeating his own novelty. Public knowledge will generally be held as prior art against the inventor, but the rule excludes privately held, proprietary knowledge from the prior art category. This arrangement, as Professor Merges observes, will encourage the inventor to invest in such knowledge prior to the beginning of an inventive project.[62] Such preliminary investigation may be useful in focusing or directing the actual project, making the conduct of the research proper more efficient.[63] Translated into the nomenclature of theories of the firm, such knowledge constitutes a project-specific asset, a type of investment that should be encouraged.

But project-specific knowledge may also provide an opportunity for hold-up if the residual rights in the knowledge are not specifically allocated. The statutory amendments condition prior art consideration of commonly developed subject matter upon whether the invention is "owned by the same person or subject to an obligation of assignment to the same person."[64] Thus the statute itself relies to some extent upon the nature of the employment agreement, and to the extent that such agreements may be incomplete, upon the courts' willingness to

---

59 See *In re Bass*, 474 F2d 1276, 1290 (CCPA 1973); *OddzOn Products, Inc v Just Toys, Inc*, 122 F3d 1396, 1401–04 (Fed Cir 1997).

60 See *In re Land*, 368 F2d 866, 879 (CCPA 1966) (holding that an inventor's solo work is not the work of another with respect to related joint work); *Shields v Halliburton Co*, 667 F2d 1232, 1235–36 (5th Cir 1982) (same).

61 See Patent Law Amendments Act of 1984 § 103, Pub L No 98-622, 98 Stat 3383, 3384–85, codified at 35 USC § 103(c); Patent and Trademark Office Efficiency Act § 4807, Pub L No 106-113, 113 Stat 1501A-572, 1501A-591 (1999), codified at 35 USC § 103(c).

62 See Merges and Duffy, *Patent Law & Policy* at 776–77 (cited in note 57).

63 Id.

64 Patent Law Amendments Act § 104, 98 Stat at 3384–85. See also Patent Efficiency Act § 4807, 113 Stat at 1501A-591.

plug gaps in the agreement. Under the conditions of employee mobility most likely to lead to hold-up, the inventions of inventive entities with overlapping memberships may frequently not be subject to assignment to the same "person." This suggests a need to fully vest firms with residual inventive rights under a doctrine of "invention made for hire."

CONCLUSION

I have suggested in this Essay that certain intellectual property doctrines may be better understood when viewed through the lens of property-based theories of the firm, and that certain doctrines, such as work made for hire, might be adjusted in light of such theories. In doing so, I have focused primarily on doctrines of ownership allocating residual rights between employees and employers. Other relevant relationships in the intellectual property context, such as partnerships or joint ventures, may also be amenable to such analysis. Similarly, we might profit by analyzing other intellectual property doctrines, such as joint authorship in copyright or double patenting, in light of property-based theories of the firm. Additionally, because different forms of intellectual property likely entail different degrees of asset specificity, firm-based comparisons between different regimes of intellectual property may be fruitful for understanding the role of intellectual property in the firm. These possibilities suggest that the theory of the firm should be a useful tool in further examination of intellectual property law.