HARVARD | BUSINESS | SCHOOL

The Impact of Modularity on Intellectual Property and Value Appropriation

Carliss Y. Baldwin
Joachim Henkel

Working Paper

12-040

November 20, 2012

The Impact of Modularity on Intellectual Property and Value Appropriation

Carliss Y. Baldwin[†]
Joachim Henkel*

December 8, 2011
Revised, November 20, 2012

[†] Harvard Business School
cbaldwin@hbs.edu

* Technische Universität München
henkel@wi.tum.de

The Impact of Modularity on Intellectual Property and Value Appropriation

Carliss Y. Baldwin,[1] Joachim Henkel[2]

November 2012

Modularity is a means of partitioning technical knowledge about a product or process. When state-sanctioned intellectual property rights are ineffective or costly to enforce, modularity can be used to hide information and thus protect intellectual property (IP). We investigate the impact of modularity on IP protection by formally modeling three different threats to the value of IP: (1) unauthorized use by known agents; imitation or substitution by third parties; and the withdrawal of IP by agents or third party owners. For each threat, we consider the impact of modularity in the presence or absence of an effective legal system. The models permit us to identify specific strategies for protecting IP and thus capturing value in modular systems. We illustrate each of the major strategies with examples from practice.

Keywords: Modularity, value appropriation, intellectual property

[1] Harvard Business School, Soldiers Field, Boston, MA 02163, USA, cbaldwin@hbs.edu.

[2] Technische Universität München, Arcisstr. 21, 80333 Munich, Germany, henkel@wi.tum.de.

The Impact of Modularity on Intellectual Property and Value Appropriation

INTRODUCTION

Distributed innovation has become increasingly important in the modern global economy. Supply chains now stretch around the world as firms outsource production to innovative suppliers (Sturgeon, 2002). At the same time, many firms have structured their products as "open systems" in which outsiders are invited to innovate (Gawer and Cusumano; 2002, Adner and Kapoor, 2010).

In general, distributed innovation is made possible by the modularity of the underlying product or process. Modularity brings many technical benefits, including the division of labor, reduced cognitive complexity, and higher adapability and evolvability (Simon, 1962; Baldwin and Clark 1997, 2000; Schilling, 2000). Yet, despite these well-known technical benefits, it is not always straightforward for firms to capture value and protect their intellectual property (IP) in a modular system. Consider the following tale of two companies.

In 1998, Valve Software released "Half-Life," a video game that was divided into two modules: the core engine and the game code (Jeppesen, 2004). Valve published the game code and granted users a broad license to modify and share it. Soon users had built a modified game called "Counter-Strike," which became far more popular than the original. To play Counter-Strike, though, players had to license the core engine from Valve, and thus Counter-Strike increased the total demand for Valve's product.

Contrast that with IBM's introduction of the PC in 1981. That product was designed as a highly modular system, and IBM outsourced almost all parts of it except an essential piece of software called the BIOS (Basic Input Output System). But just a couple years later, Compaq and Phoenix Technologies had managed to legally replicate the BIOS, enabling others to create inexpensive knockoffs that were fully compatible with the PC's hardware and software. As those cheap clones flooded the market, IBM’s PC division struggled to maintain profitability (Cringely, 1992; Ferguson and Morris, 1993).

In this paper, we will argue that the crucial difference between the two firms lies in the way they managed and were able to protect their IP in conjunction with modularity. We will show that modularity can be used to protect IP by, for example, enabling companies to disperse and hide information that might otherwise be difficult to protect through the legal system. But at the same time, modularity increases the probability of imitation or substitution by third parties unknown to the firm. Thus firms must make nuanced strategic trade-offs when using modularity to protect their IP. To capture the relevant trade-offs, we formally model three different threats to the value of IP and investigate the impact of modularity on each threat in the presence or absence of an effective legal system. The model permits us to identify specific strategies for protecting IP and thus capturing value in modular systems.

We illustrate these strategies with examples taken from practice. Several examples involve digital systems, which can be modularized at low cost and in different ways (Whitney, 2004). Yet, as we shall show, examples involving non-digital technologies exist, both in history and in modern times. Still, because digital systems are easy to modularize, as digital technologies spread, we expect the strategic use of modularity to protect IP to become increasingly important in a wide range of industries.

Our paper is organized as follows. In the next section, we review the literature. We then begin our formal analysis by arguing that modularity can be a useful tool in protecting IP when legal IP rights are imperfectly effective or costly to enforce. We go on to identify three major threats to the value of IP: unauthorized use of IP by agents of the firm, imitation or substitution by third parties, and withdrawal of IP by agents or third party owners. We use a principal-agent model to show how modularity in combination with (imperfect) legal rights can be used to address each threat, and provide examples from practice of both successful and failed strategies. We conclude the paper by describing the limitations of our analysis, implications for scholars and managers, and directions for future work.

BACKGROUND

This paper draws on three separate strands of literature: the theory of modularity, the modern theory of property rights and relational contracts, and diverse theories that explain how firms profit from innovation and IP. These literatures form the foundation of our analysis.

Modularity

According to the theory of modularity, firms can divide complex technical systems into components ("modules") that can be designed independently but function together as a whole. Three key concepts are worth noting.

First, the modular structure of a technical system is a choice that system architects make (Ulrich and Eppinger, 1994; Whitney et al., 2004). That decision is constrained by the laws of physics and the limits of the architects' knowledge, but most complex technical systems can be designed to be more or less modular, and the boundaries between modules can be located in different places (Mead and Conway, 1980; Hennessy and Patterson, 1990; Ulrich and Eppinger, 1994; Whitney et al., 2004; Baldwin, 2008; Fixson and Park, 2008).

Second, the technique of modularization involves partitioning design decisions into discrete subsets and then creating a body of design rules (also known as standards) that specify how the resultant modules will interoperate (Mead and Conway, 1980; Baldwin and Clark, 2000). If the separation of modules is done properly, the design decisions taken with respect to one module will not affect decisions taken in other modules. Design tasks can then be allocated to different organizational units or firms (Langlois and Robertson, 1992; Sanchez and Mahoney, 1996).[3]

[3] For a survey of the extensive literature of the impact of modularity on organizations and industry structure, see Colfer and Baldwin (2010).

Third, just as modules can be separated in terms of their underlying design decisions, knowledge about modules can likewise be separated. As long as they can access the design rules, Module A's designers do not need to have specific knowledge about Module B's internal structure. Thus the designers of each module have (potentially) exclusive knowledge. Conversely, designers working within a module must share knowledge or risk jeopardizing the success of their efforts. It follows that modularity is a technical means of creating non-overlapping, exclusive bodies of knowledge.

Although the technological and organizational consequences of modularity have received a great deal of scholarly attention, the strategic consequences—i.e., how modularity affects competition among firms—have not been widely studied. Notable exceptions are Rivkin (2000), Pil and Cohen (2006) and Ethiraj, Levinthal, and Roy (2008). These authors argue that modularity poses a strategic trade-off for firms: on the one hand, it makes a firm's products easier to imitate, but it also allows the focal firm to innovate faster and thus stay ahead of would-be imitators. In this paper, we generalize this prior work by looking at how modularity affects value in the presence of multiple threats to IP (in addition to the threat of imitation).

Property Rights and Relational Contracts

The economic theory of the firm is concerned with the design of incentives within companies and the location of boundaries between those organizations (Coase, 1937). Building on an earlier theory of property rights (e.g., Demsetz, 1967), Grossman and Hart (1986) and Hart and Moore (1990)—hereafter referred to as "Grossman, Hart, and Moore"—unified and extended prior theories of the firm that were based on agency and transaction costs (Williamson, 1985; Jensen and Meckling, 1986). They noted that contracts cannot be written in sufficient detail to cover all contingencies and, in any case, such legal documents are limited to specifying only behavior that a third party can verify. They then defined "property rights" as the residual rights of control over assets used in production and developed a theory of the optimal allocation of property rights.

Grossman, Hart, and Moore framed their argument in terms of the ownership of physical assets and data, such as a customer list. Brynjolfsson (1994) applied their reasoning to knowledge and intellectual property. We follow Brynjolfsson in focusing on knowledge as an asset, and we follow Hart and Moore (1990) in defining "property" as the ability to exclude others from using the asset. We differ from these prior works, however, in that we do not consider property rights to be secure. Indeed our analysis will focus on threats to IP and actions that can be taken to protect it.

Baker, Gibbons and Murphy (2002) extended Grossman, Hart, and Moore's theoretical framework to include so-called "relational contracts." In a relational contract, deviations from cooperative behavior can be punished by terminating the relationship. As long as the near-term reward to deviation is less than the long-term continuation value of the relationship, parties to the contract will cooperate without state enforcement (Greif, 1998; 2006). Relational contracts are thus said to be "self-enforcing" (Telser, 1980; Baldwin, 1983; Bull, 1987; Greif, 1998), and they can be modeled as repeated games (Bull, 1987; Baker et al., 2002), a practice that we will adopt.

Profiting from Innovation and IP

We also draw on many diverse strands of literature regarding the ways in which firms can profit from innovation and IP. From Barney (1991), we take the idea that knowledge may be a source of profits and competitive advantage, but only so long as it cannot be imitated or substituted. From Teece (1986; 2000), we take the ideas that firms must actively manage their knowledge resources and that profits may flow to the owners of complementary assets. Also from Teece (1986) and the literature on cross-country property rights (La Porta et al., 1997, 1998; Rajan and Zingales, 1995, 2001; Maskus, 2000; Zhao, 2006; Branstetter et al., 2011), we take the idea that IP rights vary by jurisdiction and may be weak or strong. Our strategy for modeling imperfect IP rights is adapted from Antràs, Desai, and Foley (2009).

From the literature on platforms and open source software (Gawer and Cusumano, 2002; Casadesus-Masanell and Ghemawat, 2006; Eisenmann, Parker and Van Alstyne, 2011a,b; Evans, Hagiu and Schmalensee, 2006; Henkel, 2006; Baldwin and Woodard, 2011; Casadesus-Masanell and Llanes, 2011), we take the concepts of “open” and “proprietary” parts of a system. In this paper we focus on proprietary systems and modules.

Finally, from the literature on markets for technology, we take the idea that there are great hurdles to setting up efficient markets for knowledge (Arrow, 1962; Arora, Fosfuri and Gambardella, 2001; Gans and Stern, 2003, 2010). In particular, we extend the analysis of Gans and Stern (2010) by modeling specific threats to IP and introducing modularity as a strategic option that can be used to protect IP.

THE VALUE OF KNOWLEDGE

Setting and Scope of the Analysis

Our first fundamental assumption is that knowledge is a source of value and economic profit. If unique knowledge is needed to create a valuable product or process, then control of that knowledge can be translated into a monopoly with a corresponding flow of monopoly rents. We define IP as knowledge that is exclusively controlled by a particular firm and thus can serve as a source of economic rent. Such property includes the classic legal forms of IP (such as patents, copyrights, and trademarks) but also confidential information that is known to the firm’s employees and suppliers and that may or may not be legally protected (Hall et. al., 2012). Consistent with the property rights literature, we consider such knowledge to be the property of a particular firm if the firm can exclude others from using it (Hart and Moore, 1990).

Our second fundamental assumption is that knowledge is divisible. Given human cognitive limitations, a key problem in designing technological systems is to divide the design tasks and related knowledge into a set of sub-problems that can be solved by specific people who share knowledge in particular ways (Simon, 1962; Parnas, 1972a,b; Clark, 1985; von Hippel, 1990). In a “one-module” system, all sub-problems are inter-related: every designer must know what the others are doing, and each must be able to share his or her knowledge and reasoning with all others. In contrast, in a modular system, the sub-problems and related knowledge are divided into independent modules, in which "every module ... is characterized by its knowledge of a design decision which it hides from all others" (Parnas, 1972b).

In essence, then, *modularity is a technical means of dividing and controlling access to design-relevant knowledge*. As such, it can be used to hide information and thus protect IP. However, if the owner of valuable knowledge already has perfect, legally enforceable *IP rights*, then he can use those rights and the powers of the state to exclude any and all others from using his knowledge. In such cases, using modularity to protect IP is unnecessary and redundant. We capture this reasoning in our first proposition:

> **Proposition 1**. *In a world of perfect, state-enforced IP rights, it is unnecessary to use technical divisions of knowledge, i.e., modularity, to protect IP.*

It follows that modularity can be a useful tool when IP rights are imperfectly effective and/or costly to enforce.

One way that modularity can be used to protect IP is by splitting crucial knowledge into separate modules. Consider the following historical example. In the eighteenth century, Frederick Augustus II, Elector of Saxony, had maintained a monopoly on European porcelain by the simple expedient of imprisoning the inventor in a fortress in Meissen. When the inventor was close to death, Augustus ordered him to divide his knowledge between two successors. One man was told the formula for porcelain paste; the other learned the secrets of making porcelain glaze. Thus, after the inventor died, no one individual could replicate the Meissen porcelain-making process (Gleeson, 1998).

But using modularity to protect IP is neither a simple nor straightforward undertaking. For one thing, as indicated, there are different types of threats to IP, and actions that increase protection against one can reduce protection against others. The major threats we will consider are (1) the unauthorized use of knowledge, particularly by a firm's own agents; (2) the imitation or substitution of knowledge by parties unknown to the firm; and (3) the withdrawal of knowledge by the firm's own agents or by outside owners of IP. In the following sections, we explain these threats in greater detail and construct a formal model to investigate the impact of modularity on each. The model will show how the threats can interact with each other and with the legal system in various non-obvious ways.

Throughout the rest of the paper, we assume that the principal has unique knowledge, which serves as the basis for a valuable monopoly. The principal has already decided which parts of the knowledge can be made widely available ("open") and which parts should remain proprietary, i.e., under his exclusive control. Our concern is with the "proprietary" parts of the system. Specifically, how can the principal use modularity to maintain exclusive control of knowledge that (in his judgement) is critical to his ability to appropriate value?[4]

THE THREAT OF UNAUTHORIZED USE OF KNOWLEDGE

When someone possesses knowledge and wants to realize its value, he must generally employ individuals and contract with suppliers who will turn that knowledge into a working product or process. But in doing so, the principal must (almost always) reveal that information to those agents, subject to the modular division of the system. Those agents, in turn, could reveal the knowledge to competitors or set up a rival establishment. This threat is well-known in law and economics and has been discussed by Teece (1986), Liebeskind (1997), Rajan and Zingales (2001), Rønde (2001), Marx (2011) and others.

One-module systems, with no enforceable property rights or contracts

We first consider the simplest case: a one-module system, in which each design decision is related to all others. Thus, people working on the module must have unrestricted access to all relevant knowledge in order to address the system's interdependencies. This leaves the principal vulnerable especially when property rights or contracts over knowledge are not enforceable within the governing

[4] Many systems contain both proprietary modules and so-called "open" modules whose IP is made public (usually with some license restrictions). Proprietary modules are sometimes called "closed," but the latter term is used ambiguously. The question of which parts of the system should be proprietary vs. open is a complicated issue in its own right, which lies beyond the scope of the present paper. We will return to it in the conclusion.

legal system. In such cases, though, the principal can still protect his monopoly by keeping the system closed to outsiders and by setting up a relational contract with his agents. Thus, our basic model is based on the concept of self-enforcing or relational contracts (Telser, 1980; Baldwin, 1983; Bull, 1987; Greif, 1998, 2006; Baker et al., 2002; Gibbons and Henderson, 2012).

Following Bull (1987) and Baker et al. (2002), we think of a relational contract between a principal and his agents as a repeated game in which the principal essentially pays the agents not to defect. For simplicity, we assume all parties are risk neutral, although this assumption is not essential to the results.

For time consistency, the principal must design the contract as a series of payments whose present value to each agent is always greater than or equal to the agent's expected payoff of defecting (Baker et al., 2002; Gibbons and Henderson, 2012). Given geometric time preference on the part of the agents, the payments can be structured as an annuity.[5] Because payments will end on dissolution of the monopoly, agents have incentives not to defect. On the other hand, if the payments stop, the agents can defect, hence the principal has incentives to continue making payments. Thus an incentive-compatible relational contract between the principal and agents is theoretically feasible.

Let the total number of agents with access to the principal's knowledge be denoted by N. The agents fall into two types. The first type, called "trustworthy," will under no circumstances defect. The second type, called "untrustworthy," will defect if it is in their own interest to do so. Each agent knows his or her own type, but not the types of the other agents. The probability that any given agent is untrustworthy is known to both the principal and all agents. We assume that untrustworthy agents decide independently whether to defect or not.[6] Apart from not knowing the other agents' types, all agents have full information about the parameters and the structure of the game.

[5] To keep notation simple, we assume that agents live forever. Assuming, instead, a constant probability of dying in each period would keep our results qualitatively unchanged.

[6] The timing of moves is as follows. Each period is divided into two sub-periods. In the first sub-period, agents simultaneously and independently decide whether to defect and the defectors leave. In the second sub-period, the principal learns if any have defected and pays the agents accordingly. The defectors, if any, collect and split their reward. Then,

Let v denote the flow of profits (rents) from the monopoly, and $V \equiv v / r$ denote the capitalized value of the rents in perpetuity. (Note: we use lower-case letters to denote cash flows, and the related upper-case letters to denote the present values. For simplicity, we assume a single discount rate r is applicable throughout.)

As indicated, agents with access to the principal's knowledge may defect to competitors. A single defector will receive a reward X that is greater than zero. In the event of defection, the principal will lose his monopoly and his establishment will also be worth X.[7] We assume that the aggregate value of the resulting duopoly is below that of the monopoly, or $0 < 2X < V$, otherwise the principal would have set up the second establishment himself. We also assume that if several agents defect at the same time, they will band together and split the reward equally, while the principal still receives X (this assumption simplifies the argument but is not essential).

To set up a relational contract, the principal promises to pay each agent a bonus above the competitive wage with a present value of Z if nobody defects and zero otherwise. The minimum bonus is affected by the principal's need to make the contract self-enforcing. Specifically, if $Z < X$ then defection becomes the dominant strategy for each agent: if all others stay, then defecting increases the respective agent's payoff from Z to $X > Z$; if n-1 other agents defect, then a switch from "stay" to "defect" increases the defector's payoff from 0 to $X/n > 0$.[8] Thus, if $Z < X$ then the unique Nash equilibrium is characterized by all agents defecting.[9]

To bring about an "all stay" equilibrium, the principal must pay every untrustworthy agent an amount whose value is equal to the maximum reward, X. And because (by assumption) the principal

conditional on no defections, the game is repeated. There is no last period of the game, although it may end probabilistically as a result of exogenous events (see below).

[7] The duopoly that results from defection may not last forever. Its expected duration is reflected in the value of X.

[8] For simplicity, we assume that only the first n agents to defect band together to set up a rival establishment, while later defectors (defecting after the principal has stopped paying them) are not allowed to join. This assumption is not essential; alternatively, we could assume that defection rewards of x/n accrue to early defectors for one period, followed by rewards of x/N in every subsequent period.

[9] The game is a (multi-player) prisoner's dilemma among the untrustworthy agents (the others never defect) if Z is larger than the payoff in an "all (untrustworthy agents) defect" situation. This will be true if the (expected) number of untrustworthy agents is relatively high.

cannot distinguish between untrustworthy and trustworthy agents, *all agents must receive a stream of payments whose value equals X.*[10] Thus the total cost of protecting the principal's knowledge against unauthorized use by agents is NX. And if the (incremental) value of the monopoly before agent payments, $V–X$, is less than NX, the monopoly is not worth protecting, and the principal will be content with X. It follows that the net value of the monopoly to the principal is the maximum of 0 and $V–(N+1)X$. We encapsulate this reasoning in our second proposition:

> **Proposition 2.** *If property rights and contracts are not enforceable, the principal cannot distinguish between trustworthy and untrustworthy agents, and the total reward to all defectors is $X > 0$, then to protect his exclusive knowledge, the principal must pay each agent an annuity worth X. The total cost of protecting the monopoly is NX, where N is the number of agents with access to the principal's knowledge. The value of the monopoly to the principal is* $\max[0, V-(N+1)X]$.

An implication of Proposition 2 is that, when X is large relative to V, the only way to sustain a monopoly is to keep N small. Interestingly, it does not help to decrease u, the fraction of untrustworthy agents, unless the principal can distinguish between who is untrustworthy and who is not.

"No Clean Sale" in the absence of enforceable IP rights

Proposition 2 has implications for the existence of markets for technology (Arrow, 1962; Arora et al., 2001; Gans and Stern, 2010). To illustrate the problem, suppose the principal, after establishing the monopoly, desires to sell it and pursue other interests. From the buyer's perspective, the principal has the knowledge and after one sale could sell it again to another party or resume the business himself. To prevent this, the buyer must include the erstwhile principal in a (new) relational contract,

[10] Strictly speaking, with $X = Z$ agents are indifferent between staying and defecting. We assume that ties are broken in favor of staying. Note also that, if the agents have a positive probability of dying, the per-period payment per agent may be greater than rX.

and pay him on an ongoing basis not to defect. Thus, if we define a "clean sale" of property as one in which the two parties do not have a continuing relationship then, under the conditions stipulated in Proposition 2, *there can be no clean sale of a knowledge-based monopoly*. In addition, the most attractive buyer, from the principal's perspective, is an existing agent because selling to an outsider increases N by 1, while selling to an agent leaves N unchanged. These points are captured in the following:

Proposition 3A. *Under the conditions set forth in Proposition 2, the principal cannot sell his exclusive knowledge (or the business it supports) without becoming himself an agent who receives ongoing payments under the relational contract.*

Proposition 3B. *Other things equal, a buyer who is already an agent of the principal can afford to pay more for the knowledge-based monopoly than an outsider can.*

The argument we are advancing is different from the Arrow Information Paradox (Arrow, 1962), which states that in the process of educating a buyer about the value of information, the seller may need to disclose that information, at which point the buyer can simply take it without paying for it. Instead, as in Anton and Yao (1994), the threat is that the seller will continue to have the knowledge after the sale; hence he must be given incentives (via a relational contract) not to sell it again.[11,12]

The Impact of Modularity

As discussed above, systems can be made modular by separating design sub-problems and hiding important information within the different modules (Parnas, 1972; Baldwin and Clark, 2000). We assume that the principal's original product or process can be divided into M modules (with

[11] In the context of a one-shot game, Anton and Yao (1994) show that a seller can use the threat of resale to elicit value from a buyer even if the buyer can costlessly expropriate the knowledge. In a multi-period model, we show that the value transferred from buyer to seller must be structured as the continuation benefit of a relational contract. In other words, for incentive compatibility, the seller must be paid each period not to defect, hence there is "no clean sale."

[12] In addition to employees and suppliers, customers may receive valuable knowledge from the principal. Gans and Stern (2010) describe a threat related to the Arrow Paradox, which they call "user reproducibility." They observe that, if users have access to cheap and accurate copying technologies, any customer may be able to re-create and transfer copies of the principal's good, thereby breaking his monopoly. In terms of our model, under conditions of (cheap) user reproducibility, N would include not only employees and suppliers of the principal but also all customers. In theory, a limited number of customers could be made party to the relational contract, but they would have to receive a stream of payments, perhaps in the form of follow-on services, equal in value to the defection reward X. In a large market, the principal's monopoly will be unsustainable, unless protected by state-sanctioned property rights.

corresponding module monopolies) that are kept closed to outsiders. For simplicity, we first consider a symmetric modularization: the original *N* agents are split evenly into *M* groups, with *N/M* agents per group. To implement a modular architecture, the principal must create *M*+1 separate bodies of knowledge: one for each module and one set of design rules spanning all modules (Mead and Conway, 1980). We assume that modularization entails an architectural cost, *A*, which is greater than zero. In addition to the upfront costs of design, *A* includes any increased coordination costs and decreased performance that may arise from splitting up decisions and restricting communication between modules (Ethiraj and Levinthal, 2004).

After modularizing the system, the principal conveys to each group of designers the design rules and the knowledge pertaining to their respective modules, and each group then works separately and independently of the others.[13]

We first consider the impact of a "value neutral" modularization, i.e., one that does not change the value *V* of the system. We make the following assumptions: (1) no module has negative incremental value; (2) the principal will always compete with a module defector, thus the principal and defector(s) will split the value of a module market between them; and (3) if all modules defect simultaneously, the agents' total defection reward is X and the principal also receives X (as above). Under these assumptions, X_m, defined as the reward to an agent who defects with particular module must be less than the reward to an agent who defects with the whole system:

$$X_m < X \quad .$$

Payments to each agent thus decline under the modular architecture, and (summing over all agents) the total cost of the relational contract declines as well.

The simplest case is that of symmetric additivity, i.e., value is evenly distributed over the modules. This case obtains when all modules are independent, i.e. neither substitutes nor

[13] This is the same network structure that is sometimes used in clandestine and revolutionary organizations. The logic behind the two designs is the same—to hide information and thus reduce the impact and rewards of defection.

complements. In this case, the reward to module defectors is *X/M*, and the cost savings to the principal from modularization is:

$$\text{Cost Savings from Modularization} \;=\; N\left[X-\frac{X}{M}\right]$$

Note that the savings increase with both the number of agents and the number of modules.

The decline in payments is even greater if the modules are functional complements, which implies that the value of the whole system is greater than the sum of the stand-alone values of all the modules (Milgrom and Roberts, 1990). In this case, the *average reward* to a module defector, $\bar{X}_m$, is less than *X/M*:

$$\bar{X}_m \equiv \frac{\sum_{m=1}^{M} X_m}{M} < \frac{X}{M} \quad , \tag{1}$$

and the cost savings from modularization are commensurately higher:

$$\text{Cost Savings from Modularization} \;=\; N(X-\bar{X}_m) > N\left[X-\frac{X}{M}\right].$$

We summarize our reasoning in the following:

Proposition 4. *A value-neutral modularization reduces the cost of protecting IP from unauthorized use by agents by reducing the average defection reward per person. This reduction is larger if the modules are complements.* [14]

We have said that modularization entails an architectural cost, denoted *A*. The principal should be willing to pay this cost if the reduction in agent payments exceeds *A*. Furthermore, despite information hiding, over time, agents working on one module might acquire critical information about

[14] The strategy of protecting trade secrets by partitioning them into separate non-overlapping subsets (modularization) has been discussed conceptually by Liebeskind (1996) and formally modeled by Rønde's (2001). Rønde uses a two-period model to show that, given weak legal protections, a firm may wish to partition (i.e., modularize) trade secrets, even if this move reduces the firm's productive efficiency (i.e., is costly). His analysis of "separate production" reaches conclusions similar to our analysis of complementary modules. However, we use relational contracts to model a multi-period game and study the modular structure of the artifact at hand, while, for the sake of simplicity, suppressing details about the product market and employer-employee contracts that lie at the heart of his analysis.

other modules. In this case, the principal would need to increase payments under the relational contract, and the cost advantage of modularity would decrease with the passage of time.[15]

The reasoning behind Proposition 4 relies on the existence of competition between the principal and module defectors. By assumption, the principal and the defector have the same knowledge, thus it is reasonable to assume that the principal is capable of producing a module at a cost comparable to the defector's. Cash flows from selling the module in turn give the principal an incentive to enter the module market. However, if for any reason (e.g. antitrust regulation), the principal cannot enter the module market, then a defector might be able to monopolize that market and earn a defection reward higher than X_m. In that case, Proposition 4 would not hold, and modularization might not reduce the cost of the relational contract.

Some modularizations increase the total value of the system. Value-increasing modularity tends to occur when one or more components have great "technical potential," defined as the capacity for improvement in response to experimentation with new designs. For example, Baldwin and Clark (2000) estimated that the modular architecture of IBM's System/360 may have increased its value by 25 times over previous non-modular computer systems. Value-enhancing modularizations can be incorporated into our analysis in a straightforward way. Let the ratio of the value of the modular system to the one-module system be denoted as α (greater than 1), and assume that defection rewards are proportional to value. Then, the average defection reward per agent will be $\alpha \overline{X}_m$. Depending on whether $\alpha \overline{X}_m$ is greater than or less than X, total payments under the relational contract may be higher or lower in the modular system than in a single-module monopoly. In either case, though, the value captured by the principal is larger in the modular than in the one-module system.

[15] We are grateful to an anonymous reviewer for pointing out this possibility. Liebeskind (1996) discusses methods and the costs of maintaining secrecy across organizational units.

The Impact of a Legal System

Up to this point, we have assumed that the principal cannot enforce his IP rights or contracts and must thus rely on self-enforcing relational contracts to prevent agents from defecting with valuable knowledge. Although worldwide IP rights have been strengthened by the recent TRIPS agreement, they are still weakly enforced in many developing countries (Kyle and McGahan, 2009; Branstetter et al., 2011). Thus we expect information-hiding modularity to be used to protect IP in such jurisdictions (see the example of R&D by multinational firms discussed below). But even in developed countries there is generally some uncertainty about the enforceability and scope of patents, copyright and trade secrets protection (Lemley and Shapiro, 2005, 2007). Thus, even in developed countries, payments under relational contracts and possibly modularity may be used to supplement state-sanctioned IP rights.

Adapting the approach of Antràs et al. (2009), we model an imperfect legal system in the following way. Let the parameter ϕ denote the weakness of IP rights to the principal's knowledge. The value of ϕ ranges between 0 and 1, depending on the surrounding legal regime and the nature of the principal's knowledge. If $\phi = 0$, the principal's IP rights are strong enough to make the rewards of defection equal to zero; if $\phi = 1$, the principal has no IP rights.[16] The defection reward then equals ϕX for the system as a whole and ϕX_m, $m = 1 \ldots M$, for an individual module.[17]

Obviously, if $\phi = 0$, then the legal system alone will be a deterrent, and (by Proposition 1) modularity will be irrelevant to the protection of IP.[18] In contrast, if ϕ is greater than 0, then agents can expect a positive reward to defection, even in the presence of state-sanctioned IP rights. In that

[16] Patents generally require the principal to make public some of his valuable knowledge. If *all* of the principal's knowledge is disclosed, then agents can expect no reward for defection and $\phi = 0$. But if some knowledge remains non-public then ϕ will be greater than zero. On partial disclosure under patents, see Hall et al. (2012).

[17] Modules may be heterogeneous with respect to their legal protection: for example, some modules may incorporate novel technology subject to patents or creative ideas subject to copyright, while other modules may use only widely available technologies and mundane ideas. Differences in the legal status of modules can be an important consideration in practice, but to simplify notation, we suppress that complexity.

[18] Mathematically, changing the cost of protecting the monopoly from $N\phi X$ to $N\phi X_m$ makes no difference if $\phi = 0$.

case, by the logic of Proposition 4, modularity can be used to reduce payments to the agents, thus increasing the value of the monopoly to the principal. It is important to note that legal systems are not only imperfect but also costly to use. We will address this issue below, after discussing the threat of imitation or substitution by third parties.

Two examples show how modularity and the legal system interact to protect valuable IP from misappropriation by employees and suppliers. The examples are based on 18th century porcelain technology and R&D projects by multinational corporations. (Other examples may be found in Liebeskind, 1997, and Henkel, Baldwin and Shih, 2012.)

Example 1—Porcelain

Earlier, we discussed how Frederick Augustus II, Elector of Saxony, used modularity to maintain a monopoly on European porcelain. Augustus had to rely on agents—chemists and artists—to carry out the porcelain-making process. This made his monopoly vulnerable because there was no perfect legal system that could effectively enforce his IP rights—a defector had only to ride as far as the nearest border (a relatively short distance) to escape his jurisdiction. In the beginning, Augustus managed to keep all the essential knowledge in the head of one man (N=1) whose movements he could control by force. He could have done the same with a single successor to that man, but instead he cleverly split the knowledge of porcelain paste and glaze between two individuals. In doing so, Augustus modularized knowledge about the porcelain-making process (M=2) in such a way that $X > X_1 + X_2$. This inequality reflects complementarity between the porcelain paste and glaze modules: glazed porcelain products were much more valuable than either unglazed porcelain or glazed pottery. Thus, even though N had increased from 1 to 2, dividing up the agents' knowledge about the process reduced the payments that Augustus had to make to prevent defection. Note that the introduction of modularity was costly—the ruler had to use soldiers to keep workers in separate zones within the fortress (Gleeson, 1998), and the division of knowledge likely entailed some coordination problems.

(There was also a risk that agents knowledgeable about different parts of the process would collude and share their knowledge, thus Augustus may have employed spies and informers as well.)

Example 2—Protecting the Value of R&D in Countries with Weak IP Rights

Zhao (2006) and Quan and Chesbrough (2010) have studied the use of modularization to protect IP in the context of multinational companies' (MNCs) research and development (R&D) across international boundaries. Because knowledge created through R&D cannot be protected effectively in countries with weak IP rights, Zhao contends that multinationals will assign projects to such countries whose results are strongly complementary with other projects conducted in the United States. She presents evidence that patents obtained by MNC subsidiaries in countries with weak IP rights have more value inside the MNC parent than outside of it. In a series of case studies and interviews, Quan and Chesbrough (2010) found that MNC managers modularized the R&D process and located projects with little stand-alone value in China because of concerns about weak IP protection in that country. In doing so, they reduced the potential rewards, or X, to defectors in those countries. The multinationals could thus take advantage of the lower cost of conducting research in developing countries and still appropriate most of the value of their R&D investments. Without such modularization, payments to prevent defection in countries with weak IP rights would have reduced (or perhaps eliminated) the cost advantage.

THE THREAT OF IMITATION OR SUBSTITUTION BY THIRD PARTIES

The second threat to the principal's value appropriation is imitation or substitution by third parties. People unknown to the principal may be able to imitate a system or module design, or create a substitute without having access to the principal's unique knowledge. Because their identity is unknown, the principal cannot include such people in any relational contract. But if imitation or substitution is likely, the value of the monopoly will decrease.

We model imitation and substitution by third parties using a hazard model, addressing the one-module case first. Let ϕs denote the probability of imitation *or* substitution in any time period. As in the previous section, ϕ measures the weakness of IP rights in the legal system and can range from 0 to 1. If ϕ=*0*, property rights are strong enough to deter all attempts at imitation or substitution.[19] The parameter *s* captures all other determinants of the probability of imitation or substitution. Consistent with our assumptions in the previous section, we assume that if imitation or substitution occurs, the principal's per-period cash flow will drop to *x* and his establishment will be worth *X*. Thus the principal obtains surplus cash flow of *v–x* as long as the monopoly endures.

Under these assumptions, the probability of the monopoly surviving from time *t* to *t+1* is $(1-\phi s)$. Using the perpetuity formula with a positive hazard rate, the value of the monopoly under this threat (after subtracting payments to agents under the relational contract, and provided $v - x - N\phi x > 0$) is:

$$q(\phi s)\cdot(v - x - N\phi x) \ , \tag{2}$$

where:

$$q(\phi s) \ \equiv \ \sum_{t=1}^{\infty}\left[\frac{1-\phi s}{1+r}\right]^{t} \ = \ \frac{1-\phi s}{r+\phi s} \ . \tag{3}$$

Equation (2) shows that the value of the monopoly can be deconstructed into two parts: (1) excess cash flows ($v - x - N\phi x$) that continue as long as the monopoly endures; and (2) a capitalization factor $q(\phi s)$ that takes into account the probability (ϕs) that the monopoly will end in any time period. Obviously, a positive probability of imitation or substitution ($\phi s > 0$) reduces the value of the monopoly to the principal. Payments under the relational contract ($N\phi X$) last only as long as the monopoly endures; thus their value goes down as well, although per-period payments remain the same.

[19] As noted above, ϕ may vary across modules. In addition, imitation and substitution are separate events that have different probabilities and likely different values of ϕ. In particular, IP rights are typically more effective against imitation than against substitution. We suppress these complexities in the interest of notational simplicity and because they do not affect the basic results.

The Impact of Modularity

Rivkin (2000), Pil and Cohen (2006) and Ethiraj et al. (2008) have argued that modularity makes imitation easier, and we agree with their logic. Basically, modularity decreases the complexity of individual components, and makes the design more transparent hence easier to imitate.

Modularity also makes substitution easier, although the mechanism is somewhat different. By definition, a modularization reduces dependencies *between* particular components and the rest of the system. So-called gateway, translator, and adapter technologies become feasible because modular interfaces are relatively simple. Designers can then focus their resources on module-level experiments, developing new designs for modular components without changing other parts (Baldwin and Clark, 2000). Ease of experimentation in turn reduces the costs and increases the likelihood of successful substitution.

Thus, all other factors being equal, modularity operates to increase *s*, the probability of imitation or substitution by third parties net of ϕ. As such, the overall impact of a given modularization must balance its effect on agent payments against the hazards of imitation and substitution. This tradeoff is most easily seen in the case of additive, symmetric modules. In that case, each module provides an equal share of system profit and has the same number of agents. Defection rewards per person are ϕX for the non-modular system and $\phi X / M$ for the modular system. Let the probability of imitation or substitution be ϕs_M for each module of the modular system. Modularization is then worthwhile if:

$$q(\phi s)\cdot[v - x - N\phi x] < q(\phi s_M)[v - x - N\phi x / M] - A \tag{4}$$

For the cases of interest where the reduction in agent payments is greater than the architectural cost, there exists a unique value, $s_M^* > s$, that equates the two sides of the expression. If the modularization results in a value of s_M less than s_M^* then it is worthwhile: otherwise it is not. However, the precise value of s_M^* depends on all other parameters in the expression.

Thus in assessing the impact of a particular modularization on his IP, the principal must trade off lower payments to agents under the relational contract against a shorter expected lifetime of the asset because of higher probabilities of imitation or substitution by third parties. Whether a particular modularization increases or decreases the total value of the monopoly depends on how these countervailing mechanisms operate module-by-module and can be assessed by aggregating the sum of value of the module monopolies. This reasoning is captured in the following:

Proposition 5. *A value-neutral modularization increases the probability of imitation or substitution by third parties and thus decreases the overall value of a monopoly to the principal. In the presence of a relational contract, however, the impact of modularization depends on the balance of countervailing effects, hence is indeterminate.*

Two examples, both involving IBM, show how modularity in conjunction with an imperfect legal system affects third parties' incentives to imitate or substitute.

Example 3—IBM PC Cloning (Threat of Imitation).

As discussed earlier, the IBM PC, introduced in 1981, was designed as a highly modular system. In order to bring it to market quickly, IBM outsourced almost all components, peripheral devices, and software but kept control of an essential module: a piece of software called the BIOS, which was protected by copyright (Ferguson and Morris, 1993, IBM, 1981).[20] However, copyrighted software can be legally imitated using a technique called "clean-room reverse engineering" (Cringely, 1992; Ferguson and Morris, 1993). (In this process, designers who have never seen the original are given detailed information about its *behavior*, and they create a new artifact that exactly mimics that behavior. The new design, by definition, is not a copy because its creators never saw the original.)

Compaq and Phoenix Technologies reverse engineered the BIOS using clean-room techniques, and Phoenix licensed its BIOS widely to clone-makers. Notably, the high degree of modularity of the

[20] We call a module "essential" if the system has a value of zero without this module or a substitute to it. Of particular interest are modules that are essential and under the control of a single party (the principal or some outside owner of IP). In contrast, a module can be essential to the system but available on a competitive market. This is the case, e.g., for the power supply module of a PC.

PC system and the small size of the the BIOS made clean-room reverse engineering possible at a reasonable cost. Had the BIOS module been larger (with commensurately more complex behavior), it would have been less vulnerable to this threat. Indeed, one expert observed:

> [The Macintosh] BIOS is very large and complex … unlike the much simpler and more easily duplicated BIOS found on PCs. The greater complexity and integration has allowed … the Mac BIOS … to escape any clean-room duplication efforts. (Mueller, 2003, p. 28)

Example 4—System/360 and Plug-Compatible Peripherals (Threat of Substitution)

IBM's System/360 was the first "truly modular" computer (Ferguson and Morris, 1993). Peripheral devices such as disk drives, tape drives, and printers could be added to an existing system without difficulty. Soon after the introduction of System/360, hundreds of new firms making peripheral devices entered the market in competition with IBM. The top managers at IBM were surprised and annoyed by this competition but were unable to prevent it (Pugh et al., 1991).

Plug-compatible firms could not legally sell pure copies of IBM equipment because such products would have violated IBM's patents, copyrights, trade secrets, and other IP rights. However, the plug-compatible devices were not merely copies but new (and often better) models that performed the same functions. Therefore they did not infringe on IBM's IP rights. Moreover, System/360's design rules—the critical interface standards—did not qualify for IP protection because they lacked novelty (Bell and Newell, 1971).

Thus, consistent with our argument, the modularity of System/360 facilitated substitution by third parties. However, in contrast to the IBM PC, the value of System/360 to IBM did not depend on controlling a single module like the BIOS. Instead, the company's profits were spread over numerous modules, many of which had no substitutes. Thus, in spite of competition from plug-compatible peripheral firms, IBM was able to capture most of the value of the System/360 (Baldwin and Clark, 2000).

Legal Protection as a Modular Option

Legal systems are not only imperfect but also costly to use. *Ex ante* expenses include the cost of acquiring property rights (such as patents) and the costs of drawing up legal documents (employment contracts, non-disclosure agreements, and licenses based on IP rights). *Ex post* expenses include the costs of monitoring, litigation, and enforcement. Let the value of expected legal costs (both present and future) be denoted by L_m, where again *m* refers to a particular module. The principal will assert IP rights over a given module only if it pays to do so. In other words, the joint deterrent impact of the legal system on third-party imitators and substitutors and potential defectors must outweigh the costs of using it.[21]

Let v_m denote the incremental value of a particular module, expressed as a per-period flow of profits. Where module values are additive, v_m equals the direct profits from the module; otherwise it represents the principal's assessment of the damage done to his overall business if he loses control of the module. Then:

Proposition 6. *For a particular module* m*, the principal will acquire imperfect state-sanctioned property rights if and only if:*

$$q\left(s_m\right)\cdot\left(v_m - N_m x_m\right) \leq q\left(\phi s_m\right)\cdot\left(v_m - N_m \phi x_m\right) - L_m \, . \tag{5}$$

(Here we assume that ties are resolved in favor of the legal system.)

To obtain legal protection, the principal may need to disclose some or all of his valuable knowledge, for example in a patent application (Hall et al., 2012). In those cases, the left-hand side represents the value of relying on secrecy to protect IP, while the right-hand side shows the value of revealing the knowledge and relying on the state for protection.

[21] For simplicity, we assume that using the legal system or not is a binary choice. In reality, the legal system may be used to varying degrees and in various ways.

Proposition 6 makes it clear that the principal's strategy for IP protection must be devised module-by-module (and in conjunction with the system's modular structure itself). In other words, there is no "one size fits all" solution.

THE THREAT OF THE WITHDRAWAL OF KNOWLEDGE

Up to this point we have assumed that the principal is the original source of all system-relevant knowledge. We now consider knowledge that the principal obtains from other sources. This threat has two forms. First, in the course of performing their jobs, agents of the principal—both employees and suppliers—may come to possess unique, valuable knowledge about the principal's products or processes. Even if their knowledge has no value outside the system (hence obtains no defection rewards), these agents may bargain for a share of system rents by threatening to withhold their knowledge. Second, in a legal system that recognizes IP rights, if the principal uses knowledge owned by someone else, the other party may demand compensation.

We model this threat in the following way. Suppose the principal is presented with a demand to share his rents or face the withdrawal of a certain body of knowledge. Consistent with the modern property rights literature (Grossman and Hart, 1986; Hart and Moore, 1990; Baker et al., 2002), we assume that the principal and the claimant will reach a settlement that is *ex post* efficient.

We consider first the simpler case of an agent's threat to deny the principal the use of her exclusive, system-relevant knowledge. Assume there is a cost, denoted Y, of operating without the agent's knowledge or, equivalently, "designing around" it. Under the assumptions of the Nash bargaining solution, to avoid this deadweight cost, the principal will pay the agent $Y/2$ and the agent will provide her knowledge to the system.[22] In this case, the cost of the threat of withdrawal, denoted W, is simply $Y/2$. Note that if the knowledge in question is essential to the system, then Y may equal

[22] The actual split depends on the relative bargaining power of the two parties, which may depend on psychological and contextual circumstances outside the scope of our analysis. By convention, the Nash bargaining solution assumes the parties have equal power in the relationship, hence split the surplus 50-50. Although details will change, our main results go through for any allocation of bargaining power, that is, any fractional split of value.

the entire value of the monopoly. The agent may then lay claim to half (or some other significant fraction) of system rents net of payments under the relational contract and for legal protection.

Next we consider a threat originating from an outside party who has legal property rights (patents, copyrights, or trademarks) to IP that the principal wants to use. Again the principal has the ability to "design around" the knowledge at a cost *Y*. But in choosing his response, he should also consider how a lawsuit might unfold. Assume that, if the external owner sues in court, she will prevail with probability θ and then be awarded damages *D*.[23] The costs of legal proceedings are *C*, which are split evenly among the parties.[24] By settling out of court, the parties can avoid this deadweight cost. The Nash bargaining solution in *this* negotiation is for the principal to pay royalties of θD to the claimant.

If $Y \leq \theta D$, then designing around the focal IP (or simply doing without it) dominates going to court. As before, the parties can avoid this deadweight cost by splitting the surplus and (under Nash bargaining assumptions), the principal will pay a royalty of *Y*/2. In contrast, if $Y > \theta D$ then designing around is not a credible threat and the principal ends up paying θD. The cost of the threat of withdrawal, *W*, then becomes:

$$W = \begin{cases} Y/2 & : Y \leq \theta D \\ \theta D & : \theta D < Y \end{cases} \tag{6}$$

Note that the parties settle in any case: Because we assume full information and zero transaction cost of negotiating, they reach an efficient outcome. Payments to the outside owner are highest when expected damages are in the intermediate range: $Y/2 < \theta D < Y$ because in this region, the threat of a design-around is not credible. Thus ironically, somewhat weaker property rights and a lower

[23] Note that θ is negatively correlated with ϕ, but the correlation does not have to be exact.

[24] This assumption is typically correct in the United States. The alternative assumption that the loser pays C is valid in many other countries. It implies that the threshold as well as the royalty in the lower line of Equation (6) (see below) become $\theta D + (\theta - 1/2)C$, but leaves our argument qualitatively unchanged.

probability of winning can benefit the outside owner because the principal will then prefer paying expected damages to designing around the IP.

The Impact of Modularity

Just as modularity can be used to reduce payments to agents under a relational contract, it can also be used to reduce payments to those threatening to withdraw their knowledge. To employ modularity for this purpose, the principal must identify where the problematic knowledge is needed in the system and encapsulate those areas in separate modules. Encapsulation reduces the cost of designing around the externally owned knowledge. (A "design around" is essentially a substitution, and we have already observed that modularity makes substitution less costly.)

We capture this reasoning in the following:

Proposition 7. *A value-neutral modularization that encapsulates knowledge controlled by others reduces the cost of designing around the knowledge, hence the cost of the threat of withdrawal.*

Sometimes, however, it is simply not possible to encapsulate external knowledge into neatly defined modules. In these circumstances agents and/or third-party owners of IP, by threatening to withdraw their knowledge or IP rights, can lay claim to a large percentage of the value of the system.

Three examples illustrate how modularity can mitigate the threat of withdrawal:

Example 5—Web Server Platform

LaMantia et al. (2008) describe a software company that sells Web-based applications. The company's entire product family depended on a single platform component containing both the company's own code and code licensed-in from another software vendor. Moreover, the platform was designed as a one-module system with many dependencies between the company's own and licensed-in code. It would have been very difficult to separate the licensed-in code from the rest on short notice, and thus the firm expected to pay a high cost to renew the license.

Anticipating this threat, the firm re-designed its platform, encapsulating the licensed-in code in a separate module. In effect, the firm prepared for designing around the licensed-in code before they faced the actual threat of withdrawal. After the modularization, the cost of substituting other code for the licensed-in module was greatly reduced. Indeed soon after the redesign, the company began to offer products that used third-party and open-source substitutes for the previously licensed-in software.

Example 6—The GPL and Proprietary Licenses

Software developers today obtain code from both commercial vendors, under proprietary licenses, and open source repositories, where the code is often governed by the so-called General Public License (GPL). Proprietary licenses generally prohibit the passing on of human-readable source code to any third party, while the GPL explicitly states that any user of a program derived from GPL code is entitled to its source code (Free Software Foundation, 1991). Thus interweaving commercially licensed code and open source code leads to conflicting legal requirements, increasing the risk of withdrawal of both types of code. These conflicts can be resolved by placing GPL and proprietary code in separate modules.[25] Through modularization, the source code of GPL modules can be revealed and the source code of proprietary modules can remain unpublished.

Our last example shows how modularity can be combined with legal rights to protect the IP of an essential module, while enabling innovation by third parties in other modules.

Example 7—Valve Software and Counter-Strike

As discussed earlier, Valve Software designed its game "Half-Life" in two parts: the core engine and the complementary game code (Jeppesen, 2004). The engine was given a proprietary license and distributed only in a machine-readable (or binary) format, while the game code was distributed as human-readable source code, and users were granted broad license to modify it. As it turned out, user-developed games, especially "Counter-Strike," became far more popular than the original game and a

[25] Provided this separation is clear enough to ensure that the proprietary code is not to be considered as "derivative work" in the sense of the GPL (Free Software Foundation, 1991).

key driver of Valve's profitability.[26] (In 2003, user-developed games accounted for 99% of Valve's monthly player hours, compared to 0.7% for in-house games and 0.3% for games produced by commercial suppliers (Jeppesen, 2004).)

Valve's insight was to see that players could be an important source of talent in the design of new games. However, they were also aware that a truly popular game, like Counter-Strike, might be viewed by players as an essential part of the system, on a par with the game engine. Thus when they published the software development kit (which was copyrighted IP), they restricted its use to non-commercial applications:

> You may use, reproduce and modify the SDK on a non-commercial basis solely to develop a modified game (a "Mod") for Valve products compatible with and using the Source Engine. … [The modified game must be] made publicly available and distributed without charge on a non-commercial basis.[27]

In this fashion, Valve avoided the need to bargain with or split system rents with another owner of IP. In a legal system with strong, enforceable property rights, Valve might have published all of its code and relied on copyright and license restrictions to protect its IP. Instead the company used a combination of modularity, secrecy, state-sanctioned property rights, and license terms. In effect, the game engine was wholly proprietary, while the SDK was partially "open." Thus even as they made parts of their IP accessible to third-party innovators, Valve prevented those agents from gaining either exclusive knowledge (game code had to be publicly available) or a profit-capturing position (non-commercial use only).

THE VALUE OF PROPRIETARY MODULES

Our analysis has focused the "proprietary" parts of a technical system in which the principal seeks exclusive control of knowledge through some combination of secrecy, relational contracts, modularity, and legal rights. As long as he protects this IP, the principal benefits from a knowledge-based monopoly and receives a stream of rents. However, the principal must protect the IP in

[26] http://planethalflife.gamespy.com/View.php?view=Articles.Detail&id=121 (accessed 10/18/12)

[27] Valve Software, Subscriber Agreement, http://www.steampowered.com/v/index.php?area=subscriber_agreement, accessed 10/18/12).

proprietary modules from misappropriation, imitation and substitution, and must pay for knowledge controlled by others. We have shown that modularity can be used to create non-overlapping knowledge sets, thus reducing the cost of protecting IP from misappropriation by agents. However, the case of the IBM BIOS clearly shows that proprietary modules are also subject to the threats of imitation or substitution if their internal structure or interactions with the system are too simple.

We can now write down an expression for the value of a proprietary module or group of modules. As before, let the subscript "*m*" denote a particular module, and let V_m^{net} denote the value of module *m* net of the cost of protecting its IP from the threats we have identified. Integrating Equations (4), (5) and (6), we have:

$$V_m^{net} = \max\left\{q\left(\phi s_m\right)\cdot\left(v_m - N_m\phi x_m\right) - L_m; q\left(s_m\right)\cdot\left(v_m - N_m x_m\right)\right\} - W_m. \qquad (7)$$

Equation (7) implies that, for each proprietary module, the principal must decide whether to seek legal protection for IP or rely on secrecy and relational contracts: this is determined by the comparison in the "max" function. From this amount, he must subtract royalty payments to agents and outside owners of IP relevant to the module: these payments are reflected in the term W_m. Finally, the probabilistic duration of the profit stream is captured by the capitalization factor, $q(\phi s_m)$.

When module values are additive, valuation using Equation (7) is straightforward. The rent stream, v_m, is simply the flow of profits (revenue minus production costs) from the sale of the module. Appropriate deductions are made for agent payments, legal expenses, and payments to outside owners of IP.

When the proprietary module is the only module that is both essential to the system and without substitutes (as in the case of the IBM BIOS), then valuation is also straightforward: v_m equals the rent stream from the entire system while deductions pertain to the module alone.

In contrast, when modules are functional complements (as in the case of Valve's core engine and games), valuation is more complicated. In such cases, v_m represents the principal's assessment of the damage done to the rent stream if he loses control of the module in question. As in all cases where there are externalities, there is no single best to way assess this value. Combinatorial value assignment methods that take account of interdependencies among subsets of modules, such as the Shapley value from cooperative game theory, can be useful in this context.

CONCLUSION

The main contribution of this paper is to provide a comprehensive answer to the question: how can a firm protect IP and thus appropriate value in a modular system? In the presence of sufficiently strong state-sanctioned IP rights, the answer to this question is trivial: the firm can simply rely on the state. However, when IP rights are imperfect—as they nearly always are—we have shown that modularity interacts with IP rights to determine a given module's vulnerability to various IP threats. The threats we considered were misappropriation of IP by agents of the original owner, imitation and substitution by third parties, and withdrawal of knowledge by agents or outside owners of IP. After systematically analyzing the impact of modularity in conjunction with the legal system on these threats, we were able to characterize the value of a proprietary module net of the cost of protecting it.

This paper makes four specific contributions that are primarily of interest to scholars. First, we defined three generic threats to the value of knowledge and showed how these could be modeled within a single framework. Second, we believe we are first to show how the threat of misappropriation of IP by a firm's agents can be mitigated by a relational contract. Third, our "no clean sale" result (Proposition 3) shows that, if IP rights are weak, the seller and buyer of a piece of IP will be bound together indefinitely in a relational contract. Fourth, we derived a formal expression (Equation 7) for the value of a proprietary module.

Our analysis also has implications for managers. First and foremost, strategies for capturing value in a modular system *must* be formulated at the module level, and the delineation of modules itself must be part of this decision. In related work, we advise practitioners to create "IP-modular" systems, in which the boundaries of modules are congruent with the firm's IP strategy (*references omitted to preserve anonymity*). "IP-modularity" in turn has two dimensions: (1) strategies for *protecting IP* in modules a firm believes it must control ("proprietary" modules); and (2) strategies for *sharing IP* in modules where the firm believes it can benefit from innovation by others (so-called "open" modules). This paper offers a comprehensive analysis of the first set of strategies, aimed at "protection." We have seen that protecting IP may require splitting some parts of the system into separate modules to hide information or to encapsulate external IP. However, elsewhere in the system, it may be desirable to combine two or more modules to make imitation and substitution more difficult. Thus managers should be aware that there is no "one size fits all" strategy or single answer to the question "how can modularity be used to protect IP?"

Our analysis has many limitations and opportunities to extend it in different directions. In the first place, there are important theoretical questions, which our analysis does not address. Specifically, we have shown how modularity and the legal system may be used to protect knowledge, but we have not addressed the question of *what* knowledge should be protected. Questions about what knowledge to share, with whom, and on what terms invite further theoretical investigation.

Another important limitation has to do with information. Our model allows for some uncertainty: for example, the principal does not know which agents are trustworthy; nor exactly when imitation or substitution will occur. For the most part, however, we assume that all parties have full information and behave rationally. Our results will still hold if the principal is uncertain about various parameters, but can form realistic expectations, adjusting them to reflect his own risk aversion. However, our model does not address the possibility of inconsistent beliefs or irrationality. In such cases, participants will make what appear to be mistakes: agents will defect with IP, imitation or substitution will arise unexpectedly, and external owners of IP will refuse to settle. Describing strategies that are robust to inconsistent beliefs and irrationality is another area worthy of theoretical investigation.

There are also a number of open empirical questions. The most promising avenue, we think, is to look within large systems to see if IP protection varies systematically across modules. Modules that are essential to the functioning of the system are of particular interest. For example, we expect essential modules that are controlled by the principal to have more IP protection than other modules.

Work on controlled essential modules may be modularized and distributed among various teams and geographies so that no single agent "knows it all." At the same time, to deter imitation and/or substitution, such essential modules will be more complex than might be dictated by purely technical considerations. Finally we expect firms to be reluctant to include IP owned by others in essential modules.

In conclusion, we would like to emphasize that modularity is not a single strategy: it is rather a large set of strategic options and related tactics that can be deployed in different ways in different places. Time and again, our theoretical analysis and empirical examples have shown there is no "one best way" to be modular: instead the best use of modularity depends on an interplay of countervailing forces. However, we hope we have convinced our readers that firms can make strategic use of modularity to protect IP and appropriate value.

ACKNOWLEDGMENTS

We are grateful to Juan Alcacer, Ron Adner, Ramon Casadesus-Masanell, Susan Cohen, Robert Gibbons, Venkat Kuppuswamy, Matt Marx, Christina Raasch, Willie Shih, Andrew Torrance, Eric van den Steen, Eric von Hippel, Jason Woodard, Dennis Yao, and Minyuan Zhao, as well as participants in seminars and workshops at the Tuck School of Business, Harvard Business School, MIT, and UNC Kenan-Flagler Business School for comments that led to significant improvements of this paper. Alden Hayashi provided excellent editorial advice. Errors and omissions are ours alone.

REFERENCES

Adner R, Kapoor R. 2010. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal* **31**: 306-333.

Anton JJ, Yao D. 1994. Expropriation and inventions: Appropriable rents in the absence of property rights. *American Economics Review* **84**(1):190-209.

Antràs P, Desai M, Foley CF. 2009. Multinational firms, FDI flows, and imperfect capital markets. *Quarterly Journal of Economics* **124**(3):1171-1219

Arora A, Fosfuri A, Gambardella A. 2001. *Markets for Technology: The Economics of Innovation and Corporate Strategy*, MIT Press, Cambridge, MA.

Arrow, KJ. 1962. Economics of welfare and the allocation of resources for invention. In: *The Rate and Direction of Inventive Activity*, (R. R. Nelson, ed.) Princeton, NJ: Princeton University Press: 609-625.

Baker G, Gibbons R, Murphy KJ. 2002. Relational contracts and the theory of the firm. *Quarterly Journal of Economics* **117**:39-84.

Baldwin CY, Clark KB. 1997. Managing in the age of modularity. *Harvard Business Review* Sept/Oct: 81-93.

Baldwin CY, Clark KB. 2000. *Design rules, Volume 1: The power of modularity.* MIT Press: Cambridge, MA.

Baldwin CY, Woodard CJ. 2011. The architecture of platforms: A unified view. In *Platforms, Markets, and Innovation*, A Gawer (ed). Edward Elgar: Cheltenham, UK: 19-44.

Baldwin CY. 1983. Productivity and labor unions: An application of the theory of self-enforcing contracts. *The Journal of Business* **56**(2):155-185.

Baldwin CY. 2008. Where do transactions come from? Modularity, transactions, and the boundaries of firms. *Industrial and Corporate Change* **17**: 155–195.

Barney JB. 1991. Firm resources and sustained competitive advantage. *Journal of Management* **17**: 99–120.

Bell, CG, Newell A. 1971. *Computer Structures: Readings and Examples,* New York, NY: McGraw-Hill.

Branstetter LG, Fisman R, Foley CF, Saggi K. 2011. Does intellectual property reform spur industrial development. *Journal of International Economics* **83**(2011): 27-36.

Brynjolfsson, E. 1994. Information assets, technology and organization. *Management Science* **40**(12): 1645-1662.

Bull, C. 1987. The existence of self-enforcing implicit contracts, *Quarterly Journal of Economics* **102** (1): 147-159.

Casadesus-Masanell R, Ghemawat P. 2006. Dynamic mixed duopoly: A model motivated by Linux vs. Windows. *Management Science* **52**(7): 1072-1084.

Casadesus-Masanell R, Llanes G. 2011. Mixed source. *Management Science* **57**(7): 1212-1230.

Clark, KB. (1985) "The Interaction of Design Hierarchies and Market Concepts in Technological Evolution," *Research Policy* 14 (5): 235-51.

Coase RH. 1937. The nature of the firm. *Economica* **4**(16): 386-405.

Colfer LJ, Baldwin CY. 2010. The mirroring hypothesis: Theory, evidence and exceptions. *Harvard Business School Working Paper* No. 10-058, January 2010 (revised June 2010), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1539592 .

Cringely RX. 1992. *Accidental Empires*, Reading, MA: Addison-Wesley.

Demsetz, H. (1967) “Toward a Theory of Property Rights,” *American Economic Review,* 57:253-257.

Eisenmann T, Parker G, Van Alstyne M. 2011a. Platform envelopment. *Strategic Management Journal* **32**(12): 1270-1285.

Eisenmann T, Parker G, Van Alstyne M. 2011b. Opening platforms: How, when and why? In *Platforms, Markets and Innovation*, Annabelle Gawer, ed., London: Edward Elgar.

Ethiraj, Sendil and Daniel Levinthal (2004) "Modularity and Innovation in Complex Systems," *Management Science*, 50(2):159-174.

Ethiraj, SK, Levinthal, D, Roy, R. (2008) "The Dual Role of Modularity: Innovation and Imitation," *Management Science* 54(5): 939-955.

Evans DS, Hagiu A, Schmalensee R. 2006. *Invisible Engines: How Software Platforms Drive Innovation and Transform Industries*, Cambridge, MA: MIT Press.

Ferguson CH, Morris CR. 1993. *Computer wars: How the West can win in a post-IBM world*. Times Books: New York.

Fixson, SK, Park, JK. (2008). "The Power of Integrality: Linkages between Product Architecture, Innovation and Industry Structure," *Research Policy* 37(8):1296-1316.

Free Software Foundation. 1991. *The GNU General Public License (GPL) – Version 2, June 1991*. http://www.opensource.org/licenses/gpl-2.0.php.

Gans JS, Stern S. 2003. The product market and the market for "ideas": Commercialization strategies for technology entrepreneurs. *Research Policy* **32**: 333-350.

Gans JS, Stern S. 2010. Is there a market for ideas? *Industrial and Corporate Change* **19**(3): 805-837.

Gawer A, Cusumano MA. 2002. *Platform leadership: How Intel, Microsoft, and Cisco drive industry innovation*. Harvard Business School Press: Boston, MA.

Gibbons R, Henderson R. 2012 Relational contracts and organizational capabilities. *Organization Science*, forthcoming.

Gleeson J. 1998. *The arcanum: The extraordinary true story*. Warner Books: New York.

Greif A. 1998. Self-enforcing political systems and economic growth: Late medieval Genoa, in *Analytic Narratives*, (ed. Bates *et al*) Princeton University Press, Princeton, NJ.

Greif, A. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, New York, NY: Cambridge University Press.

Grossman SJ, Hart OD. 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy* **94**(4): 691-719.

Hall, B, Helmers C, Rogers M, Sena V (2012) "The Choice between Formal and Informal Property Rights: A Literature Review," National Bureau of Economic Research Working Paper 17983 (April).

Hart OD, Moore J. 1990. Property rights and the nature of the firm. *Journal of Political Economy* **98**(6): 1119-1158.

Henkel J, Baldwin CY, Shih W. 2012. IP Modularity: Profiting from innovation by aligning product architecture with intellectual property. Manuscript, http://ssrn.com/abstract=2121600.

Henkel J. 2006. Selective Revealing in Open Innovation Processes: The Case of Embedded Linux, *Research Policy* **35**(7): 953-969.

Hennessy JL, Patterson DA. 1990. *Computer Architecture: A Quantitative Approach,* San Mateo, CA: Morgan Kaufmann.

IBM 1981. *IBM Personal Computer Technical Reference Manual*, Boca Raton, FL: IBM.

Jensen MC, Meckling WH. 1976. Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics* **3**(4): 305-360.

Jeppesen LB. 2004. *Profiting from innovative user communities: How firms organize the production of user modifications in the computer games industry*. Working paper No. 2004-03. Copenhagen Business School: http://ep.lib.cbs.dk/download/ISBN/ 8778690978.pdf

Kyle, Margaret and Anita McGahan (2009) "Investments in Pharmaceuticals before and after TRIPS," National Bureau of Economic Research Working Paper 15468 (October).

La Porta R, Lopez-De-Silanes F, Shleifer A, Vishny RW. 1997. Legal determinants of external finance. *Journal of Finance* **52**(3): 1131-1150.

La Porta R, Lopez-De-Silanes F, Shleifer A, Vishny RW. 1998. Law and finance. *Journal of Political Economy* **106**(6): 1113-1155.

LaMantia MJ, Cai Y, MacCormack AD, Rusnak J. 2008. Analyzing the evolution of large-scale software systems using design structure matrices and design rule theory: Two exploratory cases. *Seventh Working IEEE/IFIP Conference on Software Architecture*: 83–92.

Langlois RN, Robertson PL. 1992. Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries. *Research Policy* **21**: 297–313.

Lemley MA, Shapiro C. 2005. Probabilistic patents. *Journal of Economic Perspectives* **19**(2): 75–98.

Lemley MA, Shapiro C. 2007. Patent holdup and royalty stacking. *Texas Law Review* **85**: 1991–2049.

Liebeskind JP. 1997. Keeping organizational secrets: Protective institutional mechanisms and their costs. *Industrial and Corporate Change* **6**: 623-663.

Marx M. 2011. The firm strikes back: Non-compete agreements and the mobility of technical professionals. *American Sociological Review* **76**(5): 695-712.

Maskus KE. 2000. *International Property Rights in the Global Economy*, Washington DC: Institute for International Economics.

Mead C, Conway L. 1980. *Introduction to VLSI Systems*, Addison-Wesley, Reading, MA.

Milgrom P, Roberts J. 1990. The economics of manufacturing: Technology, strategy and organization. *American Economic Review* **80**(3): 511-28.

Mueller S. 2003. *Upgrading and Repairing PCs*, 15th ed., Indianapolis: Que Publishing.

Parnas DL. 1972a. A technique for software module specification with examples. *Communications of the ACM* **15**: 330–336.

Parnas, DL. 1972b. On the criteria to be used in decomposing systems into modules. *Communications of the ACM* **15**: 1053-58.

Pil FK, Cohen SK. 2006. Modularity: Implications for imitation, innovation, and sustained competitive advantage. *Academy of Management Review* **31**(4): 995-1011.

Pugh EW, Johnson LR, Palmer JH. 1991. *IBM's 360 and Early 370 Systems,* Cambridge, MA: MIT Press.

Quan X, Chesbrough H. 2010. Hierarchical segmentation of R&D process and intellectual property protection: Evidence from multinational R&D laboratories in China. *IEEE Transactions in Engineering Management* **57**(1): 9-21.

Rajan R, Zingales L. 1995. What do we know about capital structure? Some evidence from international data. *Journal of Finance* **50**(5): 1421-1460.

Rajan R, Zingales L. 2001. The firm as a dedicated hierarchy: A theory of the origin and growth of firms. *Quarterly Journal of Economics* **116**(1): 805-851.

Rivkin, JW. (2000) "Imitation of Complex Strategies" *Management Science* 46:824-844.

Rønde T. 2001. Trade secrets and information sharing. *Journal of Economics & Management Strategy* **10**(3), 391–417.

Sanchez R, Mahoney JT. 1996. Modularity, flexibility, and knowledge management in product and organizational design. *Strategic Management Journal,* Winter Special Issue **17**: 63–76.

Schilling MA. 2000. Toward a general modular systems theory and its application to interfirm product modularity. *Academy of Management Review* **25**: 312–334.

Simon HA. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* **106**: 467–482.

Sturgeon T. 2002. Modular production networks: A new model of industrial organization. *Industrial and Corporate Change* **11**: 451–496.

Teece DJ. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* **15**: 285–305.

Teece DJ. 2000. *Managing intellectual capital: Organizational, strategic, and policy dimensions*. Oxford University Press: Oxford, UK.

Telser LG. 1980. A Theory of Self-Enforcing Agreements. *Journal of Business* **53**(1): 27-44.

Ulrich, KT, Eppinger, SD. (1994) "Product Architecture," *Methodologies for Product Design and Development* (New York: McGraw-Hill).

von Hippel E. 1990. Task partitioning: An innovation process variable. *Research Policy* **19**: 407-18.

Whitney, Daniel E. 2004. "Physical Limits to Modularity," http://esd.mit.edu/symposium/pdfs/papers/whitney.pdf, viewed July 21, 2005.

Williamson OE. 1985. *The Economic Institutions of Capitalism*. Free Press: New York, NY.

Zhao M. 2006. Conducting R&D in countries with weak intellectual property rights protection. *Management Science* **52**: 1185–1199.

APPENDIX

Table A.1: Symbols and variables

N (N_m)	Number of agents (N_m: number of agents working on module m)
M	Number of modules
r	Discount factor
v (V)	Flow (capitalized value) of rents to the principal from the monopoly, selling the complete system; $V \equiv v/r$
x (X)	Flow (capitalized value) of rents to both the principal and a defector when both sell the complete system; $2X < V$
z (Z)	Flow (capitalized value) of bonus payments to each agent under a relational contract if the principal sells the complete system; it is shown that $Z = X$
x_m (X_m)	Flow (capitalized value) of rents to both the principal and a defector from selling module m
$\overline{x}_m$ ($\overline{X}_m$)	Flow (capitalized value) of *average* rents from selling module m
A	Architectural cost of modularizing the system
α	Ratio of value of modularized system to value of integral (one-module) system
ϕ	Weakness of IP rights; $\phi \in [0,1]$; $\phi = 0$: IP rights fully effective; $\phi = 1$: no IP rights
s (s_m)	Parameter capturing all determinants of the probability of imitation or substitution, except those related to IP rights (s_m: related to module m)
$q(\phi s)$ ($q(\phi s_m)$)	Capitalization factor; $q(\phi s) = (1 - \phi s) / (r + \phi s)$ given probability ϕs of imitation or substitution in any time period
v_m	Flow of value assigned by the principal to module m
L_m	Cost of acquiring state-sanctioned IP rights related to module m
Y	Cost of designing around exclusive knowledge owned by a third party
W	Cost of the threat of withdrawal of knowledge owned by a third party
D	Cost of legal proceedings in case the principal infringes on third-party IP
θ	Probability that the outside party prevails in court
V_m^{net}	The value of module m, net of the cost of protecting its IP against the threats of appropriation, imitation, substitution, and withdrawal of outside knowledge