The Organization of Innovation: Incomplete Contracts and the Outsourcing Decision*

Thomas Jungbauer† Sean Nicholson‡ June Pan§ Michael Waldman¶

April 23, 2021

Abstract

Why do firms outsource research and development (R&D) for some products while conducting R&D in-house for similar ones? An innovating firm risks cannibalizing its existing products. The more profitable these products, the more the firm wants to limit cannibalization. We apply this logic to the organization of R&D by introducing a novel theoretical model in which developing in-house provides the firm more control over the new product's location in product space. An empirical analysis of our testable predictions using pharmaceutical data concerning patents, patent expiration, and outsourcing at various stages of the R&D process supports our theoretical findings.
(JEL D23, L24, L65, O32)

Keywords: R&D, property rights, outsourcing, cannibalization, innovation, vertical integration

*We are grateful to Catherine Barrera, Benno Buehler, Christopher Cairns, Yi Chen, Florian Ederer, Chris Forman, Soheil Ghili, George Jakubson, Francine Lafontaine, Jorge Lemus, Sherif Nasser, Marcel Preuss, Mark Satterthwaite, Christian Schmid, Chad Syversen, Rainer Widmann, Heidi Williams, and unnamed seminar participants, in particular those of the first virtual 2020 Annual Conference of the Society for Institutional and Organizational Economics (SIOE 2020), the 2021 ZEW Annual Conference of the Mannheim Centre for Competition and Innovation (MaCCI 2021) as well as the 2021 Annual Conference of the Royal Economic Society (RES 2021) for helpful comments, and thank Kiran Butte, Yujie Feng, and Qiwei He for valuable research assistance. All remaining errors are our own. Earlier drafts of this paper were circulated under the titles "The Organization of Innovation: Theory and Evidence from the Pharmaceutical Industry," and "The Organization of Innovation: Property Rights and the Outsourcing Decision."

†Assistant Professor of Strategy and Business Economics at the Johnson Graduate School of Management at Cornell University, 401A Sage Hall, Ithaca, NY-14853, jungbauer@cornell.edu, +1(607)255-2501;

‡Professor of Policy Analysis and Management at the College of Human Ecology at Cornell University, 2307 Martha Van Rensselaer Hall, Ithaca, NY-14853, szn243@cornell.edu, +1(607)254-6498;

§VISA, Washington, DC;

¶Professor of Strategy and Business Economics at the Johnson Graduate School of Management at Cornell University, 323 Sage Hall, Ithaca, NY-14853, mw46@cornell.edu, +1(607)255-8631;

1 Introduction

A firm selling products for which its market power is substantial will want to deter the introduction of close substitutes that would cannibalize sales of these already profitable products.[1] One possibility is that this threat comes from a product introduced by a competitor.[2] Another possibility, however, is that the threat comes from new products the firm itself introduces, in which case the firm will want to control the development process in order to limit cannibalization. We show how this basic argument helps explain the organization of innovation in high tech industries. In particular, using an incomplete contracts approach related to the property rights theory of the firm (Grossman and Hart, 1986; Hart and Moore, 1990; Hart, 2017), we show both theoretically and empirically that a firm will sometimes choose in-house development over outsourcing to limit cannibalization of patented products in its existing product portfolio.

Despite the popularity of outsourcing, partnering with an outside firm is not without costs. It is well known that one of the first and foremost dangers when outsourcing tasks crucial to the success of a company is loss of managerial control (Patel, 2017). Such a lack of control is more costly if misaligned incentives affect other successful business ventures of the outsourcing firm.[3] This is particularly troublesome, as according to Lam (2004), "the needs that prompt companies to seek partners are generally ones that can't be fully specified. The product or capability may not even exist. A complete contract can't be written."

In the late 1990's and early 2000's the research-based biopharmaceutical company AstraZeneca, for example, initiated 20 new anti-cancer and 5 new biotechnology drug developments. While it outsourced the development of every single one of the biotechnology drugs, it developed all the anti-cancer drugs in-house. At the time, AstraZeneca did not hold a single patent in the class of biotechnological drugs, but had filed for 9 patents of anti-cancer drugs

[1]This basic point can be traced back to a discussion in Arrow (1962).

[2]In a recent paper, Cunningham et al. (2021) investigate competition that can lead to 'killer acquisitions,' i.e., firms with market power acquiring products in development and then discontinuing the R&D efforts.

[3]Jim Hall, president at the time of life science at advisory firm Wood Mackenzie recalls "a 60-70 million deal between a large pharmaceutical company and a medium-sized biotech firm falling apart when the biotech found that the more 'successful' the alliance became, the more money it would lose." Indeed, incompatible objectives rank among the top reasons for alliance failure in the business world (Lam, 2004).

in the previous decade, including the well-known and successful compounds "Cosudex" and "Iressa" among others.[4] In this paper, we argue that firms are in fact systematically more reluctant to relinquish control over the development, research and design of new products, the more successfully they already operate in the same product category.

We start by constructing and analyzing a theoretical model that formalizes our basic argument. In our model, a firm, called the originator, develops a new product and chooses whether to conduct the remaining R&D in-house or outsource. The firm that conducts the R&D cannot perfectly control the exact location of the new product in product space, but instead chooses a mean location and an investment level that determines the expected distance of the realized location from the mean location. A higher investment level translates into a smaller expected distance from the mean. We capture the cost of cannibalization by assuming that the originator also owns an existing patented product in the same product class as the new one.

In the model, when the originator outsources there is a contract between the originator and the firm, called the licensee, that conducts the R&D. Our focus is renegotiation proof contracts, where the contract specifies for each period who produces the product, who sells the product, and who sets the new product price. We consider a multiple periods model in which the cost of cannibalization is a function of the existing product's patent expiration date. For each possible period of patent expiration, the analysis compares the investment in location precision when the R&D is conducted in-house with the investment when it is outsourced. We then build on this comparison to characterize equilibrium behavior.

Our analysis generates three testable predictions. First, compared to firms that do not own existing patented products, a firm that does is less likely to outsource the R&D for a new product in the same product class as one of its existing patented products. Second, the probability of outsourcing is negatively related to the remaining patent duration of the existing patented product. Third, the probability of outsourcing is negatively related to the expected market share of the existing patented product at the date the new product is

[4]See the Pharmaprojects dataset.

expected to reach the market. In each case, the basic logic behind these predictions follows the same path. The incentive to outsource is lower the more profitable is the cannibalized product in the absence of the new product introduction.

We test these predictions using detailed data from the pharmaceutical industry, in which the patent system is a defining feature, and a large number of firms employ both in-house as well as outsourced development. We employ detailed compound-level data from the Pharmaprojects dataset to classify development decisions in each drug-compound year in the early part of the development process as either in-house or outsourced. We also use this dataset to construct measures concerning whether the originator of a new product owns existing patented products in the same product class. In order to control for cost-side explanations for the in-house versus outsource decision, in many of our tests we include two cost-side controls. The first is a measure of the originator's prior experience in developing drugs in the product class of the current drug under development, which could affect the outsourcing decision if learning-by-doing is important. The second measures the therapeutic scope of the originator's development portfolio, which could affect the outsourcing decision if economies of scope are important.

Our empirical analysis finds strong support for all three testable predictions. First, we find that the probability of outsourcing is lower when the originating firm has one or more existing patented compounds in the same therapeutic class as the compound under development. Second, given the presence of existing patented compounds in the same therapeutic class, the probability of outsourcing is negatively correlated with the remaining patent lengths of these compounds. Third, by supplementing the main data source with IMS sales data,[5] we show that the originator is less likely to outsource development the higher is the market share the firm expects for its same-class patented drugs at the time the new compound is expected to reach the market.

The main message of this paper is that the vertical integration decision depends on factors not covered by the mainstream theories describing boundaries of the firm. Specifically,

[5]IMS Health is an American company that provides data and related services to the healthcare industry. IMS is now part of IQVIA.

models of vertical integration typically focus either on the characteristics of an input and/or the final product(s).[6] Our argument, however, is fundamentally different. In our model, whether or not to outsource the development process depends on the characteristics of products in the firm's product portfolio for which the R&D is not an input. Even though the R&D is not an input for these "other" products, the specific nature of the R&D can affect the value of these other products resulting in a cost of outsourcing development decisions. Our empirical analysis of the pharmaceutical industry shows that this is an important driver of the in-house versus outsourcing decision in high-tech industries.

Another important aspect of our findings concerns the impact of competition on the vertical integration decision. Some of the papers that investigate the relationship between product market competition and vertical integration find a positive relationship, while others identify a negative one. We find that in the case of R&D the relationship is more nuanced. Specifically, we find that the likelihood of outsourcing increases with the level of competition. The driving force behind this result, however, is that competition negatively affects the profitability of an existing patented product. As a consequence, the incentive to avoid cannibalization of existing products by developing the new product in-house becomes less important, and outsourcing more attractive, the more competitive the market.

The outline for the paper is as follows. The next section reviews the relevant literature. Section 3 then presents the model and provides a preliminary analysis. Section 4 presents a full equilibrium analysis and discusses our testable predictions. Section 5 describes the data used in the empirical testing. Section 6 presents the empirical analysis. Section 7 discusses the extent to which our findings can be explained by alternative theories concerning the in-house versus outsource decision. Section 8 presents concluding remarks. Proofs of formal statements in the text are found in Appendix A. Appendix B generalizes our theoretical results introducing competition and multiple R&D activities.

[6]One exception is Novak and Stern (2009) that finds empirical evidence in automobile production for complementarity concerning vertical integration decisions involving inputs closely related in the production process. They provide two possible explanations for the result, neither of which, however, is related to our argument. See also Barrera and Waldman (2019) for a related analysis.

2 Literature Review

This paper contributes to the extensive literature on incomplete contracting, and, in particular, is closely related to the property rights theory of the firm which starts with the seminal analyses of Grossman and Hart (1986) and Hart and Moore (1990).[7] The property rights theory of the firm postulates that contracts are incomplete, and asset ownership grants residual control rights to the owner of the asset. In Grossman and Hart (1986), for example, two parties make non-contractible investments ex ante while utility is non-transferable. They show that one party purchases the other's assets whenever the former's investment is more important than the latter's.

We also assume that contracts are incomplete, where in our setting the incompleteness concerns development decisions involving product location, as described in the next section. When development is conducted in-house, these decisions are made by the originator, while outsourcing means they are made by the outsourcing firm or licensee. We could add an asset to our model which is owned by the originator given in-house development, and by the licensee given outsourcing. Our analysis would then be consistent with the property rights theory of the firm literature, if the owner of the asset has residual control rights concerning development decisions. We decided not to go that route, however, because we believe that in many real-world cases of outsourcing it is not ownership of an asset which is important. Rather, residual control rights are held by the firm whose employees make the decisions related to contract incompleteness. Implicitly, we assume that in-house development means decisions are made by an employee of the originator, so decisions are consistent with the interests of the originator. In contrast, with outsourcing it is employees of the outsourcing firm that make the development decisions, so decisions are consistent with the interests of the outsourcing firm.[8] Note that, as pointed out in the introduction, numerous descriptions

[7]See Gibbons (2005) and Lafontaine and Slade (2007) for surveys focused on the vertical integration decision including the property rights theory of the firm.

[8]Of course, agency problems within firms could result in decisions not being exactly in the interests of the originator in the in-house development case, and/or not being exactly in the interests of the outsourcing firm in the outsourcing case. We abstract away from this possibility in our formal analysis. However, as long as decisions are closer to the interests of the originator given in-house development, and closer to the

of outsourcing in real world markets are consistent with our assumption that outsourcing results in a loss of control for the originating firm.[9]

A related study by Aghion and Tirole (1994) employs the incomplete contracting framework to examine the organization of innovation. In their analysis, the firm's key choices are R&D efforts and financing. The results suggest that when R&D efforts are more important, R&D is more likely to be conducted by an independent unit, while financing being more important yields the opposite. Lerner and Merges (1998) study the determinants of control rights in biotechnology alliances, and find results consistent with Aghion and Tirole's theory. An important difference between our model and the Aghion and Tirole model concerns the manner in which investment outcomes depend on which party has control rights. In Aghion and Tirole (1994) (as well as Grossman and Hart (1986)), firms have different investment technologies and the firm with the superior technology purchases the assets of the other firm. In contrast, in our model all firms have access to the same investment technology.

This paper also contributes to the mostly empirical literature that relates vertical integration with product market competition. A number of early studies such as Tucker and Wilder (1977), Levy (1985), and Balakrishnan and Wernerfelt (1986) focus on US manufacturing and find a positive correlation between vertical integration and product market competition. More recently, Galdon-Sanchez et al. (2015) concerning services and Gil and Ruzzier (2018) focusing on the Spanish television industry find a negative relationship which stands in contrast to most of the earlier studies. We develop a theory that predicts a positive relationship between competition and the frequency of outsourcing, in which the positive relationship is due to reduced incentives for outsourcing when the originator owns an existing patented product in the same product class as the product under development. We also provide

interests of the outsourcing firm when development is outsourced, the qualitative nature of our results should remain unchanged.

[9]Another way in which our model differs from the standard property rights theory of the firm approach concerns the benefits and costs of vertical integration. In a standard property rights theory of the firm model, vertical integration results in inefficiencies associated with incomplete contracting. In our model, in contrast, outsourcing, i.e., not vertically integrating, results in inefficiencies due to incomplete contracting. In that sense, our model is similar to the transaction cost approach, such as Klein et al. (1978) and Williamson (1979), where vertical integration avoids inefficiencies due to hold-up.

empirical testing using pharmaceutical data that supports the predicted relationship.[10]

Several papers analyze the ability of firms to limit cannibalization when introducing a new product. Moorthy and Png (1992) show that, when consumers are impatient, a monopolist selling a product line can sometimes increase its profitability by delaying the introduction of a lower quality product which allows the firm to increase the price of a higher quality product. In Siebert (2015) a firm's optimal strategy in entering a new market in a duopoly setting with vertical differentiation is to introduce a single product. This result arises when avoiding cannibalization is more important in their setting than price discrimination. The literature on planned obsolescence in which renting is used to avoid time inconsistency concerning new product introductions such as Waldman (1993, 1996), Choi (1994) and Nahm (2004) is also closely related. We contribute to this literature by showing how the desire to limit cannibalization can affect the internal organization of the firm.

Finally, most of the prior literature on the issue of firm boundaries in the pharmaceutical industry has focused either on why firms form alliances, or the outcomes of alliances. Nicholson et al. (2005) shows that biotech companies send positive signals to investors by forming alliances with larger pharmaceutical firms, while Danzon et al. (2005) finds that success rates of complex phase II and phase III trials are higher for products developed in an alliance. Few papers focus on the characteristics of R&D projects that tend to result in in-house development rather than outsourcing. One such paper is Azoulay (2004) which finds that pharmaceutical firms are more likely to outsource data-intensive clinical trials, while knowledge-intensive trials are typically conducted in-house. Our paper is the first to offer a patent protection perspective on the choice of pharmaceutical alliance decisions at the project level.

[10] There is also a related literature concerning the correlation between product market competition and the magnitude of R&D investments. For example, Aghion et al. (2005) find a U-shaped relationship using UK manufacturing data. Our model does not directly address this issue, although we believe an extension of the model would generate a positive relationship when firms own patents in the same product class.

3 Model and Preliminary Analysis

In this section, we present a T-period model of a firm's decision to conduct R&D either in-house or through outsourcing. We then provide preliminary results concerning R&D decisions made given the firm chooses in-house development, and when it chooses to outsource. In the next section, we provide a full equilibrium analysis, and also derive testable predictions.

3.1 The Model

In our model, there is a single risk neutral firm that owns an existing patented product, and has decided to develop a new product in the same product class. We call this firm the originator. The originator exhibits a constant marginal cost c_1 for producing a unit of its existing patented product and no fixed costs. There are also generic producers that can produce the existing patented product at marginal cost c_1 and no fixed costs after patent expiration.

In addition to the originator and the generic producers, there is pool of $N \geq 2$ identical risk neutral in-licensing firms that we refer to as the licensees. The licensees command a potential cost advantage in developing the new product in comparison to the originator.[11] In particular, the originator has a fixed cost of development F which is a random draw from the probability density function $f(\cdot)$ with support (F_{min}, ∞), while the licensees incur a fixed cost of development F_L,

$$F_{min} \leq F_L < \infty. \tag{1}$$

We use Δ to refer to the difference in fixed costs, i.e., the originator's disadvantage relative to the licensees

$$\Delta = F - F_L. \tag{2}$$

We further assume that the marginal cost of production for the new product is lower for the

[11]In the case of the pharmaceutical industry, which is the industry we focus on in the empirical analysis, the cost advantage could stem from greater experience conducting randomized control clinical trials and interacting with the Food and Drug Administration, the key regulatory body in the pharmaceutical context.

firm that develops the product. Specifically, the developer has a marginal cost of production for the new product equal to c_2, while the marginal cost of production for a firm that did not develop the new product is $c_2^+ > c_2$. In other words, there are economies of scope between developing and producing the new product.[12]

We assume that there are T total periods, $T \geq 3$, with the new product being developed in period two and the patent for the new product lasting through period T. The patent on the existing product, on the other hand, expires at the end of period t_E, where t_E can take on any value between one and T. Much of our focus is how equilibrium behavior changes as a function of t_E. To simplify exposition, we assume no discounting.

If the originator chooses in-house development, it develops the new product, chooses the new product price each period, and produces and sells the new product each period.[13] If the originator chooses to outsource development, on the other hand, then there is a contract between the originator and the licensee.[14] The contract specifies for each period who produces the product, who sells the product (the firm that sells the product is the firm that receives payments from the consumers), who chooses the new product price each period, and a payment each period from the originator to the licensee which can depend on that period's new product quantity (the payment can be negative).[15]

We also assume the contract is renegotiation-proof. This assumption is described in more detail below. Furthermore, actions and outcomes associated with the development process itself are assumed to be non-verifiable and thus non-contractible. This means that, in case

[12]Results concerning the outsourcing decision would be qualitatively unchanged if we assumed no economies of scope concerning development and production, i.e., the marginal cost of production for a firm that did not develop the product is also c_2.

[13]We do not allow for a contract that would assign production, selling, and pricing decisions to another firm when the originator chooses in-house development. Given the originator is as efficient or more efficient than other firms in these activities when the originator chooses in-house development, allowing for such a contract would not change the equilibrium outcome.

[14]In our empirical analysis focused on the pharmaceutical industry, the development decisions we investigate that can be conducted in-house or outsourced are Preclinical and Phases I through III clinical trials. See Section 5 for further discussion.

[15]Implicitly, we are assuming that the new product price is non-verifiable and thus not contractible. This assumption is not essential for our main results, but rather serves to simplify the analysis. We also assume that the payment from the originator to the licensee in any period t cannot depend on the new product quantity in a different period. This assumption is also not essential for our main qualitative results, but also serves to simplify the description of equilibrium behavior.

the originator chooses in-house development, then it makes all the choices associated with the development process. But if outsourcing is chosen, then the licensee makes these choices, and payments cannot be directly contingent on these choices.[16]

Following Salop (1979), the product space is characterized by a unit circle, in which the location of the new product on the unit circle depends on the non-contractible development decisions. That is, the firm developing the new product (either originator or licensee) makes choices that serve to determine the location of the new product relative to the existing patented product. Due to the stochastic nature of the development process, however, the developer does not directly control the location of the new product but instead chooses a mean value for the location, l_M, and an investment level, k, that determines the expected absolute distance between the mean location and the realized location.

To be precise, the clockwise distance between the new product and the existing patented product on the unit circle is given by

$$l = l_M + \varepsilon, \tag{3}$$

where ε is a random draw from one of the following two uniform distributions:

$$U[-\alpha, \alpha] \text{ and } U[-\beta, \beta], \quad \alpha < \beta < \frac{1}{2} - (c_2 - c_1).$$ [17]

A higher investment level translates into a higher probability the random draw is from the uniform distribution with the smaller range. Let $p(k)$ denote the probability that ε is drawn from $U[-\alpha, \alpha]$ given the investment level equals k. We assume that $p(\cdot)$ is continuously differentiable, $p(0) = 0$, $p'(0) = \infty$, $p'(k) > 0$ and $p''(k) < 0$ for all $k \geq 0$, and $p(\infty) < 1$.

Due to the importance of our assumption that development decisions are non-contractible, it is worthwhile providing some additional discussion concerning this assumption. In our

[16] See Casas-Arce et al. (2019) for a recent related analysis in which details of the development process are left incomplete in a setting in which renegotiation is possible. These authors point out that this approach is consistent with commonly observed contracting practices in R&D intensive industries.

[17] The assumption $\beta < \frac{1}{2} - (c_2 - c_1)$ is imposed to ensure that some consumers on both 'sides' of the new product's location purchase the new product at every t, $2 < t \leq T$.

model, there are two development decisions—the mean location for the new product and the investment in location precision. Although we formally assume that both decisions are non-contractible, our analysis only depends on the latter being non-contractible. That is, because there is no conflict concerning the mean between the originator and the licensee, results would be unaffected if we instead assumed that the mean location was contractible and only the investment in location precision was not.

Implicitly, our assumption is that it is easier for a firm to ensure that the employees whose decisions affect the new product location provide efficient effort when development is conducted in-house rather than outsourced. We believe this is a natural assumption. When development is in-house, the originator has direct control over compensation and other incentives (such as promotion decisions) and can use those to elicit optimal effort levels. When development is outsourced, the originator does not have this type of direct control over employee incentives, and it is difficult to write a contract at a level of detail such that the outsourcing firm elicits efficient effort from the relevant employees. As indicated earlier, this basic argument is consistent with various descriptions of outsourcing in real-world markets.

On the demand side, there is a continuum of consumers of unit mass uniformly distributed along the circumference of the circle. A consumer can buy any weakly positive number of one of the products, i.e., a consumer can buy units of the originator's existing patented product or units of the new product, but we do not allow mixing.[18] To be precise, the valuation a consumer places on consuming unit q of a product is given by

$$V(q) = V^{+} - vq, \tag{4}$$

so the valuation function is characterized by decreasing marginal utility of consumption. A consumer also faces a distance cost for consuming a product not at the consumer's exact location in product space. The distance cost a consumer incurs from consuming a unit

[18] The assumption that there is no mixing is consistent with typical demand behavior in the pharmaceutical industry. For example, individuals treated for depression seldom take different anti-depressant drugs at the same time due to concerns of possible unwanted drug interactions.

located a distance s from the consumer's location in product space equals ds, $d > 0$. We also assume V^+ to be sufficiently large such that the market is always fully covered in equilibrium. For any product price, P, a consumer who chooses to buy that product purchases the amount that maximizes net utility from consumption, i.e., the consumer chooses the value for Q that maximizes

$$\int_0^Q (V^+ - vq)dq - (P + ds)Q.$$

In turn, in facing prices for the two products, a consumer chooses to purchase the product that results in the highest net utility for the consumer given the quantity choices that maximize net utility.[19]

The timing of the game is as follows. At the beginning of the first period, the originator chooses a price for the existing patented product for the first period, consumers make purchase decisions, and the value for F is realized and publicly revealed.[20] In the first period, the originator also decides whether to develop the new product in-house or outsource development to a licensee. If the originator chooses to outsource, then the first period proceeds with a contracting stage. In particular, each firm in the pool of licensees makes a take-it or leave-it offer of a licensing contract to the originator and the originator chooses a licensee.[21]

At the beginning of the second period, the originator chooses a second period price for the existing patented product and consumers again make purchase decisions. If the patent has not expired, then the originator sets the monopoly price, while it if has expired then competition with generic producers means the price equals marginal cost c_1. The developer (either originator or licensee) also chooses a mean value for the new product's location in product space and an investment level in location precision. These choices are the private

[19]An alternative theoretical approach is to assume that each consumer purchases zero or one unit of the product that provides the consumer with the highest net utility, but that there are multiple consumers in each location whose valuations are uniformly distributed over $[0, V^+]$. This alternative specification is mathematically equivalent to the specification we analyze.

[20]The assumption that the realization of F is publicly revealed is not essential for our results. In particular, results would be the same without this assumption if we focused on Perfect Bayesian equilibrium rather than Subgame Perfect Nash equilibrium.

[21]We assume that a licensee that works as the developer of the originator's new product does not develop any other product. We thus abstract away from the possibility of cross-subsidization which is the focus of Lerner and Malmendier (2010).

information of the developer. After these choices have been made, the noise term is drawn from the respective uniform distribution. Thus, by the end of the second period the new product's location in product space is determined. We assume this location to be publicly observable but not verifiable by the courts.

In the third period, the new product is brought to the market. In case the patent on the existing product expires before the third period, then the price for the existing product is at marginal cost, due to Bertrand competition, and the firm with control rights for the pricing of the new product takes this price as given when choosing a price for the new product.[22] If the patent on the existing patented product has not expired, then the originator chooses prices for both products if it has control rights for the pricing of the new product. On the other hand, if the patent on the existing patented product has not expired and the licensee has the control rights, then the two prices are determined by Bertrand competition between the two firms. In the following periods, if any, prices are determined using the same rules as in the third period. Also, our focus throughout is Subgame Perfect Nash equilibrium.

As mentioned above, we restrict our analysis to contracts that are renegotiation-proof.[23] This means that the contract between the originator and the licensee must be such that, in every period starting with period two, the parties do not have an incentive to renegotiate. To be specific, at the beginning of each period starting with period two, if the originator (licensee) makes a take-it or leave-it offer of a new contract to the licensee (originator), the equilibrium contract is such that the licensee (originator) has at least a weak incentive to turn down any offer that would make the originator (licensee) better off. That is, the equilibrium contract is such that no Pareto-improving renegotiation outcome exists.

Finally, in this paper we assume that R&D investments associated with developing the new product are successful with certainty. This is clearly a simplification, since in real world settings R&D investments frequently fail. Introducing a probability of failure, however,

[22]If we assumed something other than Bertrand competition, such as for example Cournot competition, the existing product's price would be above marginal cost. But that would not change our qualitative findings concerning the originator's outsourcing decision.

[23]Focusing on renegotiation-proof contracts is a standard approach employed in many contracting papers. For early papers that focus on how the possibility of renegotiation affects equilibrium contracting, see, for example, Dewatripont (1988), Hart and Moore (1988), and Demougin (1989).

would complicate the model without changing the qualitative nature of the equilibrium. So, in order to make the logic of our analysis more transparent, we assume the R&D investment is always successful.[24]

3.2 Preliminary Results

We start with results concerning the nature of the equilibrium contract when the originator chooses to outsource. As captured in Lemma 1 below, production, sales, and pricing are all assigned to the licensee in every period after patent expiration of the existing product. In contrast, prior to patent expiration of the existing product, production is assigned to the licensee, but sales and control rights for pricing remain in the hands of the originator.

Lemma 1 *Consider an equilibrium to the game in which the originator has chosen to outsource development of the new product. Then, the contract between the originator and the licensee satisfies i) through iii).*

i) In any period t, $2 < t \leq t_E$, the contract assigns production to the licensee, but sales and control rights for pricing of the new product remain with the originator. Also, the payment from the originator to the licensee is a fixed amount plus the number of new units sold that period multiplied by c_1.

ii) In any period t, $t > \max\{2, t_E\}$, the contract assigns production, sales, and pricing of the new product to the licensee. Also, the payment from the originator to the licensee is a fixed amount.

iii) The fixed payments from the originator to the licensee sum to the fixed amount that results in zero expected profits for the licensee.

The logic for part i) is as follows. First, the licensee is assigned production of the new product, because of economies of scope between development and production. Second, in

[24] Also, an implicit assumption, which is important for our analysis, is that the originator cannot sell the patent for the original product to an outsourcing firm. One reason such a strategy might not be feasible is the existence of private information on the part of the originator concerning the value of related products, and the resulting adverse selection problem, as in Akerlof (1970), which prohibits the possibility of trade.

any period prior to patent expiration of the existing patented product, the joint profits of the originator and licensee are maximized by giving sales and control rights over pricing to the originator, so that it can choose prices that maximize the joint profits of the two products. Given the contract is renegotiation-proof, this means that sales and control rights for the pricing decision must be assigned to the originator. Also, having the payment from the originator to the licensee be a fixed amount plus the number of new units sold multiplied by c_2 means higher joint profits, because in choosing prices the originator will internalize all the returns associated with the pricing decisions.

Now consider a period after patent expiration of the existing product, i.e., part ii) of the lemma. If the contract assigns production, sales, and pricing to the licensee, the contract will not be renegotiated because the licensee can set the price just as effectively as the originator after patent expiration. In turn, since assigning sales and the pricing decision to the licensee increases the licensee's investment incentives, this is the equilibrium outcome. Also, the payment from the originator to the licensee is a fixed amount, so that the licensee internalizes all of the effects of its pricing decision. Finally, iii) follows given competition across licensees ensures a zero expected profit condition on the part of the licensee.

The next step is to consider decisions concerning new product location. Before proceeding, however, it is worth noting that results below are to an extent driven by the optimal location of the new product. Both before and after patent expiration, profits from sales of the two products are maximized when the new product is located a distance exactly $\frac{1}{2}$ from the original patented product. This is not due to a business stealing effect since we assume parameters are such that the market is fully covered. Rather, locating the second product a distance exactly $\frac{1}{2}$ from the original patented product minimizes the average distance of consumers from the nearest product, which translates into higher profits because of an increase in average willingness to pay.

Let $L(j, t_E)$ denote the mean distance between the new product and the existing patented product as a function of whether development is in-house, $j = I$, or outsourced, $j = O$, and the period of patent expiration, $t_E = 2, ..., T$. Similarly, $K(j, t_E)$ is the investment in location precision as a function of whether development is in-house or outsourced and the period of patent expiration.

Lemma 2 *Holding all parameters fixed, if the in-house versus outsource decsion is taken as given, then i) through v) describe $L(j,t_E)$ and $K(j,t_E)$.*

i) $L(I,t_E) = L(O,t_E) = \frac{1}{2}$ *for all* t_E, $t_E = 2,...,T$.

ii) $K(I,1) = K(I,2) = K(O,1) = K(O,2)$.

iii) $K(I,t_E) > K(O,t_E)$ *for all* $t_E > 2$ *and* $K(O,T) = 0$.

iv) $K(I,T) > K(I,T-1) > ... > K(I,2) = K(I,1)$.

v) $K(O,1) = K(O,2) > K(O,3) > ... > K(O,T) = 0$.

Consider first what happens when the originator chooses in-house development. As discussed, for any value of t_E profits are maximized when the new product's location is exactly half way around the unit circle from the location of the existing patented product. So $L(I,t_E) = \frac{1}{2}$ for all t_E, $t_E = 1,2,...,T$. Now consider the investment in location precision. The firm's return to having the new product's location close to the mean location is higher prior to patent expiration, because the firm is choosing two prices. So the investment level increases the later is patent expiration of the existing patented product, i.e.,

$$K(I,T) > K(I,T-1) > ... > K(I,2) = K(I,1). \tag{5}$$

Suppose the originator opts for outsourcing and $t_E = T$. In this case, in each of periods 3 through T which are after the new product's location has been determined, joint surplus is maximized if the originator receives the returns associated with new product sales and has control rights over the pricing decision. So that is what is specified in the contract given our focus on renegotiation-proof contracts. In turn, this means that the licensee has no incentive to invest in location precision, so $L(I,T) = L(O,T) = \frac{1}{2}$ and $K(I,T) > K(O,T) = 0$. Note that the mean location specified in the contact is still $\frac{1}{2}$ as it maximizes joint surplus.

Finally, suppose that the originator chooses outsourcing and $2 < t_E < T$. Because the contract must be renegotiation-proof and the patent is still valid for sales of the existing patented product up through period t_E, sales and control rights for pricing the new product reside with the originator up through t_E. In contrast, after period t_E, the patent has expired with the result that sales and control rights for pricing the new product reside with the licensee. The result is that the licensee's incentive to invest is higher than when $t_E = T$, but lower than when $t_E = 1$ or 2, and in this range the incentive to invest falls with t_E, i.e.,

$$K(O,1) = K(O,2) > K(O,3) > ... > K(O,T) = 0. \tag{6}$$

Also, the incentive to invest is less than under in-house development, i.e., $K(O,t_E) < K(I,t_E)$ given $2 < t_E < T$, since with in-house development the developer in every period sells the product, has pricing control rights, and therefore internalizes all the returns associated with the pricing decisions. Lastly, similar to the other cases, $L(O,t_E) = \frac{1}{2}$ given $2 < t_E < T$.

4 Equilibrium Analysis and Testable Predictions

This section starts with a characterization of the in-house versus outsourcing decision and then presents our testable predictions.

4.1 Equilibrium Analysis

The focus of our analysis is the originator's choice concerning whether to conduct development in-house or to choose outsourcing. The potential benefit to outsourcing is that the fixed cost of development is lower by the amount Δ. The cost of outsourcing is that, as shown in the previous section, the expected investment in location precision is lower if $t_E > 2$, and this serves to lower originator profits because the expected distance in product space between the new product and the existing patented product is smaller. A comparison of this benefit and cost determines whether the originator chooses in-house development or outsourcing.

In the analysis that follows, our focus is the originator's choice of in-house development versus outsourcing as a function of the difference in fixed costs associated with in-house development. Proposition 1 below captures that there is a critical value for this difference, call it Δ^*, $\Delta^* \geq 0$, such that the originator chooses in-house development when $\Delta \leq \Delta^*$ and outsourcing otherwise.[25] The straightforward logic for this result is that in-house development is chosen when this choice is associated with a cost advantage or small disadvantage, while outsourcing is chosen when there is a large disadvantage associated with in-house development. The proposition also captures additional results which we discuss below.

Proposition 1 *Holding all other parameters fixed, there exists a value Δ^* such that the originator chooses in-house development when $\Delta \leq \Delta^*$ and chooses outsourcing otherwise, where Δ^* is a strictly increasing function of t_E for $t_E \geq 2$ and equals 0 if $t_E = 1$ or 2. Also, equilibrium behavior satisfies results in Lemmas 1 and 2, where the equilibrium contract given $\Delta > \Delta^*$ is unique up to the timing of the payments described in iii) of Lemma 1.*

Consider first $t_E = 1$ or 2. In these cases, the patent on the existing product expires before the new product reaches the market. As found in the previous section, when this is the case there is no disadvantage in terms of investments in location precision from choosing outsourcing. This means the investment in location precision is independent of whether the originator chooses in-house development or outsources. As a consequence, this choice depends solely on which of the two options has lower fixed costs associated with the development process, i.e., $\Delta^* = 0$ in this case.

Now consider what happens when $t_E > 2$. From the analysis in the previous section, we know that for these parameterizations a licensee's incentive to invest in location precision is less than the originator's incentive to invest given in-house development. Given this advantage associated with in-house development, the originator will only choose outsourcing if there is a sufficiently large reduction in the fixed cost of development associated with outsourcing, i.e., $\Delta^* > 0$ if $t_E > 2$.

[25] To simplify the exposition, we assume that the originator chooses in-house development whenever it is indifferent between the two choices.

Finally, consider two values for t_E, t' and $t'+1$, where $t' \geq 2$. From the analysis in the previous section, we know that the investment in location precision increases with t_E given in-house development, while it decreases with t_E given outsourcing. In other words, the expected loss due to a lower investment in location precision when the originator chooses outsourcing is higher when $t_E = t'+1$. It follows that the reduction in fixed costs associated with outsourcing required for the originator to make that choice must be higher when $t_E = t'+1$, i.e., Δ^* increases in t_E.

4.2 Testable predictions

We now discuss testable predictions of our theory. The first two testable predictions follow immediately from results stated in Proposition 1.

Testable Prediction 1 *A firm developing a new product has a lower probability of choosing outsourcing if it sells an existing patented product in the same product class, and the new product is expected to reach the market before this patent expires.*

Testable Prediction 1 follows from Proposition 1, i.e., $\Delta^* = 0$ given $t_E = 1$ or 2 and $\Delta^* > 0$ for all $t_E > 2$. Recall that Δ^* determines the probability that outsourcing is chosen, with a higher value for Δ^* translating into a lower probability that the choice is to outsource. Proposition 1 says that when $t_E = 1$ or 2, i.e., at the time the new product reaches the market the patent on the existing product will have expired, that $\Delta^* = 0$. In other words, in this case the in-house versus outsource decision is determined solely by which choice results in lower costs. If $t_E > 2$, however, i.e., the patent on the existing patented product will be valid at the date the new product reaches the market, then $\Delta^* > 0$ which means that outsourcing is only chosen if it is associated with a meaningful cost advantage.

Testable Prediction 2 *Consider a firm developing a new product that owns an existing patented product in the same product class. The longer this patent is expected to be valid after the new product reaches the market, the lower is the probability the firm chooses to outsource.*

Testable Prediction 2 is the Proposition 1 result that Δ^* increases with an increase in t_E. As before, Δ^* determines the probability of outsourcing. A higher value for Δ^* means a lower probability of outsourcing. The proposition states that an increase in t_E increases Δ^*, with the underlying logic being that an increase in t_E raises investment incentives given in-house development, but does not given outsourcing. Thus, the underinvestment given outsourcing rises with an increase in t_E, which means the fixed cost reduction associated with outsourcing needed for outsourcing to be chosen is higher. This is equivalent to saying that when the patent on the existing patented product is expected to be valid for a longer period of time after the new product reaches the market, i.e., t_E increases, Δ^* rises which translates into a lower probability of outsourcing.

The third prediction concerns market share. In our base model, the originator is a monopolist in the product class. Suppose that, however, rather than being a monopolist, the originator was one of a small number of firms selling products in the product class. In this case, the return to the originator of increased location precision would be positively related to the market share of the firm's existing patented products at the date the new product would reach the market. If this share was low, then cannibalization would be mostly in terms of other firms' patented products and sales of products not under patent protection, so the firm's incentive to control product location of the new product would be low. But if the share was high, then the firm's incentive to control product location of the new product would be high because the return to avoiding cannibalization of its own patented products would be high.

In Appendix B we formalize this logic by extending our formal analysis to a setting with a rival that sells a competing product in the same product class as the originator's products. In particular, in this extension the product space is an equilateral triangle. Consumers are located around the triangle, while the originator has an existing patented product located at one vertex of the triangle, while the rival has its competing product located at a different vertex. The originator introduces a new product and increased investments in location precision decrease the absolute distance of the realized location from the optimal location.

We formally show that, because of investment incentives concerning product location, the probability the originator chooses outsourcing decreases with the originator's market share prior to the new product introduction (see Proposition 2 in Appendix B).

Below we translate this result into a testable prediction.

Testable Prediction 3 *Consider a firm developing a new product that sells existing patented products for which the patents are scheduled to expire after the new product reaches the market. The probability of outsourcing will be lower the larger is the predicted market share of the firm's existing patented products at the date of the new product's introduction.*

One way to think about this prediction is to focus on the two returns to location precision in our argument. One return is that by reducing the expected deviation between the realized location of the new product in product space and the optimal location, the firm increases the profitability of the new product. The second is that by reducing this expected deviation, it also increases the profitability of its existing patented products prior to the expiration of those patents. Increasing the market share of the existing patented products in the product class makes the second factor more important, which means an increase in the returns to improved product location. When that market share is higher, we expect a lower probability of outsourced development, since outsourcing decreases investments in location precision.

A final point concerning testable predictions concerns the relationship between the model and the data. In the model, there is a single in-house versus outsource development decision for each product. In the data, in contrast, products are associated with multiple R&D investments, each of which can be conducted either in-house or outsourced. With this in mind, in Appendix B we extend the model to allow for multiple R&D investments for a single product and show that the testable predictions hold for each development decision in this extension of our analysis.[26]

[26]To be precise, in this extension we show that our first two testable predictions hold for each R&D investment in our multiple investment extension (see Proposition 3 in Appendix B). Although we do not include such an analysis, we could also formally show that the third prediction is also true by combining our first extension concerning incorporating a rival with our second extension concerning multiple R&D investments.

5 Data from the Pharmaceutical Industry

The pharmaceutical industry is in many ways an excellent setting in which to test our theory for a number of reasons. First, the industry spends a substantial amount on R&D for the development of new drugs each year. Second, the industry heavily relies on patents, i.e., it is possible to measure variation in the cost of cannibalization. Third, uncertainty in the development process implies that the pharmaceutical industry is a prime example of an environment in which firms cannot perfectly control the exact location of new products in product space. For example, the types of patients recruited for clinical trials, and how well the drug works on those patients will affect the FDA-approved label, and pharmaceutical firms can only market 'on-label.' Fourth, it is common practice at pharmaceutical firms to develop some new drugs in-house, while outsourcing the development of others. This suggests sufficient variability concerning the in-house versus outsource decision to test our theory.[27]

5.1 Main Data Source and Descriptive Statistics

Our principal data source is the Pharmaprojects dataset. This dataset was assembled by the company Informa and contains information concerning the development of new pharmaceutical projects throughout the world. The dataset covers information for the time period 1989 through 2004. For each chemical compound under development, the dataset contains the name of its originator, the therapeutic class, active ingredients, patent number and its filing date, if any, whether development was outsourced and, if so, the names of the licensees, and the beginning and end dates of licenses and development stages.

A key issue for our empirical analysis is defining whether development decisions are in-house versus outsourced. In particular, whether a development decision is in-house or outsourced depends on whether the originator ever signs a licensing contract and, if it did,

[27]See Lakdawalla (2018) for a survey of the literature on the pharmaceutical industry focused primarily on innovation, pricing, and marketing decisions.

the stage of the development process at which the earliest development decision was signed.[28] Clearly, if there was never a licensing contract, then all development decisions are in-house which is how we categorize them. On the other hand, if there was a license at some point in the development process, we categorize decisions made prior to the license as in-house, and those made after the first license as outsourced.[29] Below we provide additional detail concerning this classification process.

Table 1 presents a summary of the development phases as described by the FDA. Preclinical consists mostly of tests on laboratory animals, while phase I focuses on safety and phase II on effectiveness and side effects. Both phase I and phase II are typically conducted on a relatively small scale, with the former recruiting around 20 to 80 subjects and the latter between a few dozen and about 300 subjects. Phase III continues testing on safety and effectiveness employing a much larger sample, usually ranging from several hundred to $3,000$. Conceivably, a developer could affect a new drug's location in product space during phase III by recruiting specific patient population groups and testing for specific efficacy measures and side effects. However, in most cases, once phase II is completed and the FDA meets with the developer to discuss plans for phase III, it becomes quite difficult for the developer to make significant changes that would affect the new drug's expected location in product space.

With this in mind, in our main analysis we assume all development decisions occur prior to the beginning of phase III. More specifically, our focus is drug-compound years prior to the beginning of phase III, where drug-compound years that are prior to the first license are

[28]Note that when two firms merge, our main dataset updates the company name of the originator of the compound to the name of the acquiring firm. For example, if Warner-Lambert was the originator of compound X in 1997 and was acquired by Pfizer in 1999, then in our main dataset it is possible that WarnerLambert would not be identified as the originator which could create misclassification if there was a licensing contract between Warner-Lamber and Pfizer prior to 1999. We compiled a list of mergers and acquisitions and assigned compounds to the correct originating firms to avoid any statistical problems related to such misclassification.

[29]It is standard in the pharmaceutical industry for substantial control rights to be allocated to the licensee when a contract is signed. See Lerner and Merges (1998) for a discussion. Thus, it seems likely that subsequent decisions will be consistent with outsourcing rather than in-house development, which matches our choice to classify all drug-compound years after the first license as outsourced development rather than in-house development.

classified as in-house observations, while drug-compound years that are the year of the first license or later are classified as outsourced. Restricting focus to compounds for which there was at least one drug development license, the initial license occurred during pre-clinical testing in 51.4 percent of the cases, during phase I in 7.9 percent of the cases, during phase II in 12.8 percent of the cases, during phase III in 13.6 percent of the cases, and in 14.6 percent of the cases the first license was only agreed to after the product was launched. See Figure 1 for a graphic illustration.

We construct three measures of whether the originator of a new product owns existing patented products in the same product class.[30] Note that, in our main empirical analysis, we employ the broad therapeutic classes listed in footnote 30 in defining whether the originator has an existing patent in the same product class as the new product being developed. In Subsection 6.4 where we present robustness checks, we reproduce our main tests employing a set of narrower therapeutic classifications. The first measure, which we call $EOP1$, is an indicator variable that takes on a value of one if the originator owns one or more other patented products in the same product class and a value of zero otherwise. The second measure, which we call $EOP2$, equals the number of other patented products owned by the originator which are in the same product class. With a third measure, called $EOP3$, we distinguish between existing patents in later development stages from those in earlier stages. Our goal is to construct a measure of the expected number of other patented compounds in the same product class owned by the originator that will eventually reach the market. Following Higgins and Rodriguez (2006), we construct $EOP3$ using a count of same-class same-firm patented compounds weighted by the probability of becoming an approved drug conditional on the current stage of development. Based on existing research, the probabilities are 0.08 for pre-clinical, 0.20 for phase I, 0.28 for phase II, 0.58 for phase III, and 1.0 for launched drugs.

[30]According to the Pharmaprojects Therapeutic Class Codes, there are 17 broadly defined categories. These categories are alimentary/metabolic products, blood and clotting products, cardiovascular products, dermatological products, formulations, genitourinary products, hormonal products, immunological products, anti-infective products, anticancer products, musculoskeletal products, neurological products, anti-parasitic products, respiratory products, sensory products, biotechnology products, and miscellaneous products.

We also employ patent length variables. The first patent length variable, called $LOP1$, is the remaining length of the patent with the largest remaining length of all the patents owned by the same firm which are in the same class as the drug under development. The second patent length variable, $LOP2$, is the sum of the remaining patent lengths of all the drugs in the same product class owned by the same firm as the drug under development. The third patent length variable, $LOP3$, is the weighted sum of the remaining patent lengths for all the drugs in the same product class owned by the same firm as the drug under development, where the weights are based on the probabilities that drugs still under development reach the market as a function of the current stage of development.

One potential alternative explanation for how an originator chooses between in-house development and outsourcing is that it chooses to develop a drug in-house when it has a relatively large amount of experience.[31] In some of our empirical tests, we thus control for an originator's prior experience in developing drugs in the same product class as the drug under development. We define class-specific experience as the sum of the compound-year observations within a firm's particular therapeutic class up to the current year. Including this variable in our tests allows us to separate the cost-side explanation for the in-house versus outsource decision from our explanation which concerns investment incentives.

In addition to class-specific experience, in some of our tests we also include a measure of economies of scope as another cost-side control. Following Danzon et al. (2005), we construct a Herfindahl Hirschman Index (HHI) for each firm's therapeutic scope by summing the squares of the percentage of compounds being developed for each therapeutic class within a firm in a given year. The bigger the value of the HHI, the more concentrated is the firm's development portfolio in terms of therapeutic class. One might hypothesize that economies of scope arise when the development portfolios are less concentrated, and so a lower HHI should be associated with a higher probability of in-house development.

We also construct two variables that capture market-level variability in a firm's incentive to avoid cannibalization. The first variable, PDM, is the number of patented drugs on the

[31]See, for example, Pisano (1990) for empirical evidence that experience is an important determinant of whether firms choose in-house development or outsourcing.

market that are in the same therapeutic class as the drug under development, but which are owned by firms other than the originator of the drug under development. The second variable, TDM, is the total number of drugs on the market that are in the same therapeutic class as the drug under development. The purpose of including these two variables is to control for the possibility that an increase in the number of competing drugs on the market, holding fixed the total number of drugs on the market, lowers the incentive for the originator to avoid cannibalization. Table 2 provides a full list of our constructed variables along with their definitions.

Table 3 reports descriptive statistics for the development decision and patent data. The sample used in the main analysis contains $64,505$ compounds originated between 1989 and 2004. On the compound-year level, 81.0 percent of the observations are classified as in-house development. The data for the first patent existence measure, $EOP1$, shows that in 74.9 percent of the compound-year observations characterized by in-house development and 59.4 percent of the observations characterized by outsourcing, the originator owned at least one other patented compound in the same product class. In addition, compared to compound-year observations characterized by outsourcing, in-house observations have a higher number of other patented compounds in the same product class owned by the originators ($EOP2$), as well as a higher expected number of other patented compounds owned by the originators that are in the same product class and that are expected to reach the market ($EOP3$).

Table 4 shows how in-house development varies across the therapeutic classes in our dataset. In most cases the percentage of observations classified as in-house development is reasonably close to the percentage in the full sample. However, for the antiparasitic therapeutic area in-house development is substantially more common than in the overall sample, while in-house development is substantially less common in the biotechnology area than in the overall sample.

5.2 Secondary Data and Descriptive Statistics

To supplement the analysis, we also use the IMS dataset to create a market share measure based on drug sales. The IMS data includes a list of all drugs and their annual sales in the US between 1992 and 2004. The sales data are merged into the principal dataset from Pharmaprojects based on the name of the drug, its therapeutic class, as well as whether the drug is branded or not (i.e., generic). Because the IMS dataset classifies drugs according to the Anatomical Therapeutic Chemical (ATC) system which differs somewhat from the classification system used in the Pharmaprojects dataset, the tests that employ our market share measure only focus on the drug classes that are common to both classification systems.[32] We calculate for each drug under development in the principal dataset the summation of market shares for other patented drugs in the same class in each year. This variable, referred to as MSP, allows us to test how market share based on sales data affects a firm's incentive to choose in-house development. Table 5 provides descriptive statistics for the control variables.

6 Empirical Tests

In this section we empirically test the theoretical predictions derived in Section 4. We start with the two predictions concerning the role of other patents owned by the originator, and then consider the prediction involving market share. Robustness checks are presented at the end of the empirical analysis.[33]

6.1 Patent Existence

To investigate whether the data is consistent with the first testable prediction which concerns patent existence, we estimate the following Logit specification:

[32]Of the drugs that are launched in the Pharmaprojects dataset, we are able to match 57 percent to a listing in the IMS dataset. The 12 classes of drugs common to both datasets are alimentary/metabolic products, blood and clotting products, cardiovascular products, dermatological products, genitourinary products, hormonal products, anti-infective products, musculoskeletal products, neurological products, anti-parasitic products, respiratory products, and sensory products.

[33]Additional related tests can be found in Pan (2016).

$$Prob(Y_{ijkt} = 1) = \Lambda\left(\alpha_0 + \alpha_1 EOP_{ijkt} + \alpha_2 X_{ijt} + \alpha_3 Z_{jkt} + \alpha_4 W_{jt} + C_k + T_t\right). \quad (7)$$

$\Lambda(\cdot)$ is the standard logistic CDF. The subscripts i, j, k, and t index compound, firm, therapeutic class, and year, respectively. Y_{ijkt} is an indicator variable for in-house development. EOP_{ijkt} is a patent existence variable, where in some specifications it is the indicator measure $EOP1$, in other tests it is the indicator measure $EOP2$, and in others it is the measure of the expected number of other patented products that will eventually reach the market $EOP3$. X_{ijt} is a vector of development phase indicator variables. Note that, as discussed earlier, only years prior to phase III are included in our tests, because after the beginning of phase III decisions are highly unlikely to affect the degree of cannibalization. Hence, in the analysis pre-clinical testing is the omitted comparison group, and controls for phase I and phase II trials are included. Z_{jkt} is the originator's development experience in the therapeutic class of the drug under development, while W_{jt} is the originator's therapeutic scope.

Equation (7) also includes therapeutic class fixed effects, C_k, to control for unobserved class characteristics that affect both patent existence and development integration decisions. Year fixed effects, T_t, control for across-time differences in firms' preferences concerning the in-house versus outsource decision. From Testable Prediction 1 we expect α_1 to be positive, i.e., there is a predicted positive correlation between owning an existing patented product in the same product class as the product under development and choosing in-house development.[34] Table 6 reports the regression results. Each patent existence measure ($EOP1$, $EOP2$, and $EOP3$) is estimated under two different specifications. In the first specification therapeutic experience and therapeutic scope are omitted, whereas they are included in the second specification. All regressions employ robust standard errors to account for

[34]We do not include firm fixed effects because, in most cases, a firm has a very limited number of products under development in a product class, so including firm fixed effects would eliminate significant variation we use to estimate the coefficients of interest.

heteroskedacticity. Also, standard errors are clustered at the compound level to account for potential correlation concerning the in-house versus outsource decision for a particular compound across observations.

The first two columns report results for $EOP1$. The coefficient on $EOP1$ is positive and statistically significant at the one-percent level in each regression. The coefficient on the experience variable is also positive and statistically significant at the one-percent level, while the coefficient on the scope variable is negative and statistically significant at the one-percent level. The former result is consistent with therapeutic experience lowering the costs of in-house development and thus making in-house development more likely, while the latter result is consistent with therapeutic scope lowering costs which makes in-house development more likely (recall that our scope variable is such that a higher value means a less diversified portfolio of projects).

Columns 3 and 4 consider the same set of tests focusing on our patent count variable, and columns 5 and 6 consider the same tests focusing on the variable that measures the expected number of other patented products in the same class that will eventually reach the market. In each set of tests the results are similar to those in columns 1 and 2. The only difference worth noting is that in the column 4 specification the coefficient on the experience variable is positive but statistically insignificant.

Our preferred specification is the one which controls for experience and scope. Using this specification, we now report results concerning how much less licensing occurs prior to phase III when the originator owns patented products in the same product class. The baseline probability that the development of a drug will be observed to be outlicensed by the originator in a year prior to phase III is 19.0 percent. Employing the coefficient on $EOP1$ in column 2, the probability of licensing prior to phase III is 5.0 percentage points lower when the originator owns at least one other patented compound in the same therapeutic class relative to the probability when the originator owns no such patented compound. We can also do similar calculations employing analogous coefficients in columns 4 and 6. Using the coefficient on $EOP2$ in column 4, we estimate that an increase in the number of other patented compounds in the same class owned by the originators equal to one standard deviation is associated with a decrease in the probability of outsourced development by 7.7 percentage points in each year prior to phase III. Using the coefficient on $EOP3$ in column 6, we estimate that an increase by a standard deviation of the expected number of other same-class compounds owned by the originators that are expected to reach the market is associated with a decrease in the probability of outsourced development equal to 2.4 percentage points.

6.2 Patent Length

To investigate whether the length of patents owned by originators in the same product class as the drug under development is positively correlated with the originator's decision to develop the product in-house, we estimate the following Logit specification:

$$Prob(Y_{ijkt} = 1) = \Lambda\left(B_0 + B_1 LOP_{ijkt} + B_2 X_{ijt} + B_3 Z_{jkt} + B_4 W_{jt} + C_k + T_t\right). \tag{8}$$

LOP_{ijkt} is a patent length variable, where in some tests it is the length of the longest patent of the drugs the originator owns in the same class as the drug under development $LOP1$, in other specifications it is the sum of the patent lengths of the drugs the originator owns in the same class as the drug under development $LOP2$, and in yet other specifications it is the weighted (by phase-specific success probabilities) sum of the patent lengths of the drugs the originator owns in the same class as the drug under development $LOP3$. The control variables for development phase X_{ijt}, therapeutic experience Z_{jkt}, scope W_{jt} and fixed effects for therapeutic category C_k and year T_t are the same as in Equation (7). From Testable Prediction 2, we expect B_1 to be positive, i.e., an originator with longer patent life for drugs it owns in the same product class as the drug under development should be more likely to choose in-house development.

Table 7 reproduces the regressions in Table 6 where we substitute our patent length variables for the patent existence variables. We start by discussing the results in columns 1 and 2 which employ the patent length variable $LOP1$. The results here are similar to what we saw for patent existence in Table 6. First, the coefficient on $LOP1$ is positive in both regressions, and statistically significant at the one-percent level in each regression. Second, in column 2 the coefficient on the experience variable is positive and statistically significant at the one-percent level, while the coefficient on the scope variable is negative and statistically significant at the one-percent level. The results in columns 3 and 4 for $LOP2$ and in columns 5 and 6 for $LOP3$ exhibit similar patterns, except that in column 6 the coefficient on $LOP3$ is positive but not statistically significant, while the coefficient on the experience variable in column 4 is also positive but not statistically significant.[35]

As in the case of patent existence, our preferred specification is the one in which experience and scope variables are included. So we focus on that specification in reporting results concerning the magnitude of the correlation between the patent length of patents of other products in the same class owned by originators and the probability development is conducted in-house. Consider the coefficient in column 2 on $LOP1$. This coefficient tells us that a one standard deviation increase in the length of the longest patent held by the originator in the same product class as the product under development is associated with a 4.0 percentage point decrease in the probability a compound will be in a licensing contract in a year prior to phase III relative to the base level.

We can also conduct similar exercises using results from column 4. Employing the coefficient on $LOP2$ in column 4 yields that an increase in the sum of patent lengths of patented drugs in the same product class owned by the originators of one standard deviation is associated with a decrease in the probability of outsourced development by 7.6 percentage points relative to the base level.

[35]As will become apparent, specifications that include $LOP3$ are in general less supportive of the theory than those using the patent existence variables or the other patent length variables. There are strong positive correlations between the various patent variables and the experience variable, and we suspect this might be affecting the statistical significance of the coefficients on $LOP3$ in the specifications that include that variable.

6.3 Market Share

This subsection considers tests related to Testable Prediction 3, which is that outsourcing should be less common when the originator's market share in the product class is larger. We conduct two sets of tests related to this prediction. The first uses the Pharmaprojects dataset to consider how the number of competing drugs owned by firms other than the originator affects the originator's incentive to choose in-house development or outsourcing. In the second we use IMS sales data to construct an expected market share measure for an originator, and then directly test how expected market share affects the choice of in-house versus outsourced development.

As indicated, we start with tests concerning the number of drugs in the product class owned by other firms. In the following Logit specification, we develop a test by interacting the patent existence measure, EOP, with the number of other firms' patented drugs on the market that are in the same therapeutic class as the drug under development. The specific equation that we estimate takes the following form:

$$\begin{aligned} Prob(Y_{ijkt} = 1) = \Lambda(\gamma_0 + & \gamma_1 EOP_{ijkt} + \gamma_2 EOP_{ijkt} PDM_{jkt} + \gamma_3 PDM_{jkt} \\ & + \gamma_4 TDM_{kt} + \gamma_5 X_{ijt} + \gamma_6 Z_{jkt} + \gamma_7 W_{jt} + C_k + T_t). \end{aligned} \tag{9}$$

TDM is the total number of drugs on the market that are in the same therapeutic class as the drug under development, while PDM is the number of patented drugs on the market in the same therapeutic class as the drug under development but owned by different firms than the originator of the compound of interest. Given that we control for TDM, the theory predicts that the correlation between in-house development and the existence of a patent owned by the originator that is in the same product class as the product under development should be smaller when there is a higher number of competing patented products on the market owned by other firms, i.e., γ_2 is predicted to be negative. Note that other controls are the same as in Equations (7) and (8) and our focus is our preferred specification which includes controls for experience and scope.

Table 8 reports results for estimating Equation (9). Column 1 shows results when $EOP1$ is the patent existence variable. The main result here is that the coefficient on the patent existence variable is positive and statistically significant at the one-percent level, while the coefficient on the interaction term is negative and statistically significant at the one-percent level. Columns 2 and 3 show results for the same test, except that in column 2 $EOP2$ is the patent existence variable, while in column 3 it is $EOP3$. The results in these tests are similar but somewhat weaker than in the column 1 test. In particular, in column 2 the coefficient on the interaction term is negative but not statistically significant, while in column 3 the coefficient is negative and statistically significant at the five-percent level. Overall, the results in this table support the prediction.

We now consider a similar set of tests, except our focus is the effect of patent length rather than patent existence on the in-house versus outsourcing decision. In particular, we estimate a Logit specification similar to Equation (9), except now the explanatory variable of interest is a measure of patent length rather than patent existence. We again focus on our preferred specification which includes controls for experience and scope.

Results are reported in Table 9. Column 1 reports results where $LOP1$ is the patent length measure. In this column the coefficient on $LOP1$ is positive and statistically significant at the one-percent level, while the coefficient on the interaction term is negative and statistically significant at the one-percent level. Columns 2 and 3 consider the same tests as in column 1, except that in column 2 LOP2 is the patent length variable and in column 3 it is LOP3. The pattern of results in columns 2 and 3 is the same as in column 1. Overall, we again find results consistent with the third testable prediction.

We now consider a second approach for measuring how competition from other firms' patented products in the same product class affects the correlations we found in the previous subsections. In particular, rather than focusing on the number of other patented products owned by other firms, we focus on expected market share of the originator's existing patented products in the same product class as the product under development. Note that construction of our market share measures requires IMS data which only covers the years between 1992 and 2004. Here we only look at sales data for the 12 therapeutic classes listed in footnote 33. For both reasons the sample sizes for these tests are smaller than for previous tests.

According to Testable Prediction 3, the expected market share when the new drug reaches the market should be negatively correlated with the probability of outsourced development. That is, it is not the market share at the time of a development decision which should matter, but rather the expectation at the time of a development decision concerning the market share that the firm will have once the new product is introduced. Note that for purposes of this test, we assume that the firm has perfect foresight concerning future market shares since we have no data that measures market share expectations at the time of a development decision.

In Table 10 we examine the correlation between market share and in-house development. The top panel of the table reports results for the Logit specification in Equation (10):

$$\begin{aligned} Prob(Y_{ijkt} = 1) = \Lambda(\zeta_0 + &\zeta_1 MSP_{ijkt} + \zeta_2 EOP_{ijkt} \\ &+ \zeta_3 X_{ijt} + \zeta_4 Z_{jkt} + \zeta_5 W_{jt} + C_k + T_t). \end{aligned} \tag{10}$$

For each drug under development belonging to an originating firm j and therapeutic class k, MSP_{ijkt} is the market share based on the current year-t sales of patented drugs for the same firm and class. The other regressors are defined the same way as in Equation (7). In the bottom panel, we estimate a similar equation except the current MSP measure is replaced with projected future MSP. The specific equation estimated is given in Equation (11):

$$\begin{aligned} Prob(Y_{ijkt} = 1) = \Lambda(\eta_0 + &\eta_1 MSP_{ijkt+\tau} + \eta_2 EOP_{ijkt} \\ &+ \eta_3 X_{ijt} + \eta_4 Z_{jkt} + \eta_5 W_{jt} + C_k + T_t). \end{aligned} \tag{11}$$

.

As indicated, the value for MSP in this specification is the expected value at the current date of what MSP will be at the date the new product is introduced given successful development. Thus, the number of years in the future the expectation concerns depends on the development phase of the observation. We base this value on DiMasi et al. (2003) which estimates the average time a drug spends in each development phase.[36]

The top panel shows that the current year MSP is negatively correlated with originators choosing in-house development decisions, but the effect is not statistically significant. In contrast, in the bottom panel we find that expected future MSP is positive and statistically significant at the five-percent level in all specifications. We also find that in both the top and bottom panels the coefficient on the patent existence variable is always positive, and it is statistically significant at the one-percent level in five of the six regressions, where the coefficient on $EOP3$ in the bottom panel is positive but not statistically significant. These findings are consistent with Testable Prediction 3.

In Table 11 we rerun the tests in Table 10 but replace the patent existence variables with the patent length variables. The results are similar. In the top panel the coefficient on current MSP is negative but never statistically significant, while in the bottom panel the coefficient on expected future MSP is always positive and statistically significant at the five-percent level. Also, in both the top panel and bottom panel the coefficients on the patent length variables are always positive, where the coefficient is statistically significant at the one-percent level except for the coefficient on $LOP3$ in the bottom panel which is not statistically significant.

Note further that the results concerning expected future MSP in Tables 10 and 11 suggest that the effect of expected future MSP on in-house development is economically as well as statistically significant. That is, the various coefficients on expected future MSP in the bottom panels of Tables 10 and 11 indicate that a one percentage point increase in expected future MSP is associated with an increase in the probability that a drug will be developed in-house in a given year by between 4.3 and 4.9 percentage points.

[36]Based on DiMasi et al. (2003), for observations in the pre-clinical phase the expectation is ten years after the observation, for observations in phase I it is also ten years, and for phase II it is eight years.

6.4 Robustness Checks

In this subsection we consider the robustness of our results in two respects. We first consider whether results are robust to how we categorize whether an originator chooses in-house or outsourced development. We then consider whether results are robust to how we define therapeutic categories. In the analysis above observations are drug-compound years prior to the start of phase III trials, since after the beginning of phase III decisions should not affect the degree of cannibalization. One might argue, however, that the design and nature of a drug is mostly fixed as early as the completion of phase I testing. With this in mind, in Tables 12 and 13 we rerun tests reported in Tables 6 and 7 with the single change that observations are drug-compound years prior to the beginning of phase II trials.

Table 12 reports results using our patent existence variables and our preferred specification. The results are similar to the results in Table 6. The coefficient on the patent existence measure is positive in each regression and also statistically significant at the one-percent level in each regression. Table 13 reports results for our patent length variables. Here the coefficient on the patent length variable is positive and statistically significant at the one-percent level when the patent length variable is $LOP1$ (column 1) and $LOP2$ (column 2), while in column 3 the relevant coefficient is positive but not statistically significant.

As indicated, our second set of robustness tests concerns the way we define therapeutic categories. In particular, one might be concerned that our therapeutic categories are too coarse to accurately capture the cannibalization effect that our theory focuses on. That is, if a firm is currently developing one drug that we classify as being in the same therapeutic class as another patented drug the firm owns, but in reality the two drugs are in different markets, then choosing to develop the new drug in-house will not be be due to the firm's incentive to limit cannibalization and protect the value of the patent on the other drug.

To address this concern, we redefine our main explanatory variables, i.e., patent existence and patent length of other drugs owned by originators in the same therapeutic class, by using a set of narrower therapeutic classifications. For example, whereas before anti-arrhythmic drugs and cardiostimulant drugs were classified as being in the same therapeutic class, now they are in separate classes. Results for this modified specification are reported in Table 14. The coefficients on the patent existence and patent length variables are all positive and statistically significant at the one-percent level, except the coefficient on $LOP3$ in column 6 which is positive and statistically significant at the five-percent level.

7 Alternative Explanations

In this paper, we provide a novel theory of why firms choose to outsource research and development for some new products while they choose to conduct R&D for other similar products in-house. Two major alternative explanations for this phenomenon are the following. First, in many instances the decision between in-house development and outsourcing depends on a trade-off between providing incentives for research effort and minimizing finance costs. The basic argument, put forth initially in Aghion and Tirole (1994), is that an integrated structure is chosen when providing incentives for research effort is the more important concern, and vice versa. Note that this theoretical approach differs substantially from ours. Their focus is the probability of successful development, while our argument concerns the new product's location in product space, and how that might affect the value of existing products through cannibalization.

While we do not doubt that the perspective developed by Aghion and Tirole (1994) is an important factor in many real-world integration decisions concerning research and development, we feel that their logic is an unlikely explanation for our findings. According to that theory, firms with existing successful patents should be less financially constrained. Therefore, consistent with our findings, a firm with an existing patent should be more likely to choose in-house development as financing costs are less of a concern. However, this alternative argument does not explain why current market share of existing patented drugs in the same product class owned by the originators is less successful in predicting in-house development than is future expected market share, as shown in Tables 10 and 11. Moreover, it does not account for why patents in the same product category as the product under development should be particularly important for the in-house versus outsource decision. Overall, our argument about limiting cannibalization does a better job of explaining our empirical findings than a tradeoff between financing and research concerns.

Another potential explanation for the in-house versus outsource decision when it comes to R&D is a learning curve argument. Firms may choose to develop some products in-house because of lower costs associated with learning-by-doing. That is, prior R&D investments in the same product category as the product under development may reduce the costs of conducting the R&D in-house. Even though in our empirical analysis we control for both experience and scope, one might still suspect that the correlation we find between our patent existence and length variables and in-house development to some extent reflects past specific investments. This argument, however, fails to explain why the positive correlations between in-house development and the various patent existence and patent length variables are weaker when the market is crowded with competing patented drugs owned by other firms, as found in a number of our tests. Furthermore, the specific investment argument also does not explain our empirical findings concerning expected market share of other patented drugs owned by the originator. That is, if the main driver of our results is a learning-by-doing argument, it is unclear why in Table 10 there is a strong positive correlation between in-house development and expected market share at the time the new product is anticipated to reach the market, but no significant correlation with current market share.

In addition, there are a number of other alternative theories for the in-house versus outsource decision, none of which appears to be a good match for our empirical findings however. One argument, for example, is that learning-by-doing on the part of licensees can be important. In particular, an originator may choose to outsource because a potential licensee has significant experience with developing products in the relevant product category, and due to learning-by-doing it is the low-cost developer. This theory neither explains our main findings concerning patent existence and patent length, nor why expected future market share rather than current market share is positively correlated with choosing in-house development though. Another argument, put forth initially in Azoulay (2004), is that data intensive R&D activities are more likely outsourced, while knowledge intensive R&D activities are typically conducted in-house. This argument also does not explain our results concerning patent existence and patent length, or our market share findings. In summary, to the best of our knowledge the literature does not provide an alternative theory of the in-house versus outsource decision that accounts for our empirical findings as well as our model of limiting cannibalization does.

8 Conclusion

This paper focuses on the idea that limiting cannibalization of existing patented products owned by originators is important for understanding a firm's decision concerning whether to develop new products in-house or outsource. We begin by constructing and analyzing a theoretical model in which ownership of existing patented products in the same product class as a new product decreases the incentive for an originator to outsource development. The logic is that a licensee has less of an incentive than the originator to avoid cannibalizing the value of current patented products owned by the originators, so outsourcing is disadvantageous when avoiding such cannibalization is important. The model generates testable predictions concerning outsourcing, patent existence, patent length, and market share for existing patented products.

We employ data from the pharmaceutical industry to investigate the model's predictions. Our empirical findings are consistent with the theoretical predictions. First, controlling for firm characteristics and therapeutic class, we find that an originator with existing patented products as the product under development is less likely to outsource development. Second, the probability of outsourcing also decreases with the patent length of patented drugs owned by the originator that are in the same therapeutic class as the drug under development. Third, the correlation between in-house development and our patent existence and patent length variables is weaker when there are more same class patented drugs owned by other firms, holding fixed the total number of same class patented drugs on the market. Fourth, using market share data based on drug sales, we find that the probability development of a new drug is outsourced is negatively correlated with the originating firm's expected future market share of its existing same class patented drugs at the date the new drug is expected to reach the market. Also, our results are robust to alternative specifications of outsourcing and therapeutic classes.

These findings suggest that avoiding cannibalization of existing products is an important factor in determining whether a new product is developed in-house or outsourced. We focus on the incentive for in-house development rather than outsourcing when the originator owns existing patented products in the same product class, and the originator wants to control the design of the new product. A complementary perspective is that in-house development is also important when the originator owns existing patented products in the same product class that are about to expire, and as a result it is important for the originator to control the timing of the new product introduction. We feel this is an interesting topic for future research.[37]

Another interesting direction for future research is extending the analysis in terms of the nature of the innovation process. In this paper, we focus on a producer developing a new product of given quality that potentially cannibalizes sales of the firm's existing patented products. It would be interesting to investigate what possible additional insights might arise if the nature of the investment such as the quality of the innovation was endogenously determined, and also how rivalry in the innovation process itself might affect our basic conclusions.[38]

[37] See Williams (2013) for a related analysis focused on how current intellectual property protection affects the rate of subsequent innovation.

[38] See Gans and Stern (2000) and Igami (2017) for recent analyses of rivalry in the innovation process, and Budish et al. (2015) and Krieger et al. (2017) for recent analyses focused on the nature of investments such as the quality of innovations.

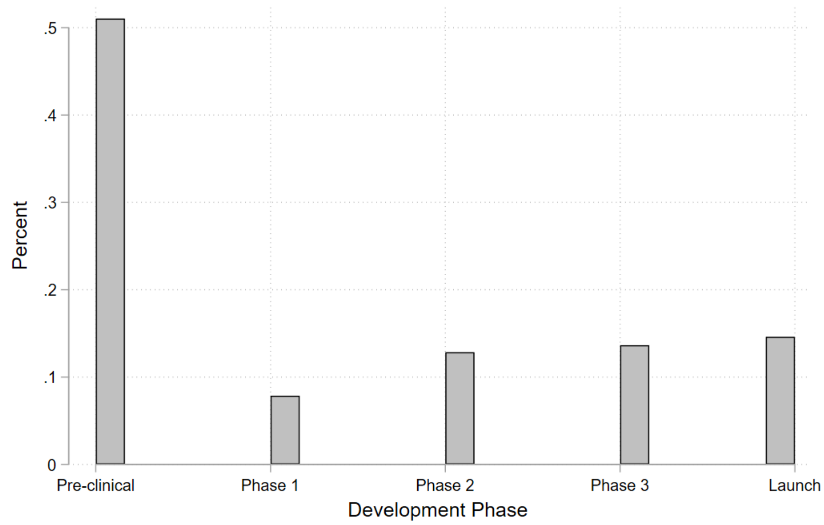

Figure 1: Distribution of the Earliest Licensing Deal by Development Phase

Table 1: Summary of Development Phases

Drug Development Stage	Description (according to the FDA)
Pre-clinical trial	Submission of investigational new drug application for the FDA to review. Companies need to show results of pre-clinical testing on laboratory animals and propose plans for human testing;
Phase I trial	Usually conducted in healthy volunteers to determine the most frequent side effects, as well as how the drug is metabolized and excreted. Number of subjects range from 20 to 80. Emphasis is on safety;
Phase II trail	Obtain preliminary data on whether the drug treats a certain disease or condition. Number of subjects range from a few dozen to about 300. Continues to evaluate safety and short term side effects;
Phase III trial	The FDA and the sponsors meet to determine how large-scale studies in Phase III should be done. Gather more information on safety and effectiveness. Studies different populations, dosages and combined usage of other drugs. Number of subjects ranges from several hundred to about 3,000 people;

Source: http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm

Table 2: Definiton of Constructed Variables

Variable	Description
In-house	Indicator equals 1 if compound is never licensed out by the originating firm or if its earliest licensing deal was made after the start of Phase III trials
Existence of Patents	
EOP I	Indicator equals 1 if at least one other compound in the same therapeutic class and same firm is patented
EOP2	Number of other patented compounds in the same therapeutic class and same firm
EOP3	Expected number of other patented compounds in the same therapeutic class and same firm that would reach the market
Length of Patents	
LOP1	Length of the longest patent among other compounds in the same therapeutic class and same firm
LOP2	Sum of the patent lengths among other compounds in the same therapeutic class and same firm
LOP3	Weighted sum of patent lengths among other compounds in the same therapeutic class and same firm according to their probability of becoming an approved drug
Other	
Experience	Cumulative count of compound-year observations within a firm for a therapeutic class corresponding to the compound of interest
Scope	Sum of the squares of the percentage of compounds being developed for each therapeutic class within a firm in a given year
PDM	Number of patented drugs on the market in the same therapeutic class but not the same firm as the compound of interest
TDM	Total number of drugs on the market in the same therapeutic class as the compound of interest
MSP	Market share based on sales for existing patented drugs in the same class and same firm as the compound of interest

Table 3: Descriptive Statistics on In-house Development and Patent Profile

Number of compounds					11,402
Number of firms					577
Years covered					1989-2004

	Outsourced Compounds Mean (SD)	In-house Compounds Mean (SD)	Overall Mean (SD)	Min	Max
	Level of Observation: compound-year (64,505)				
In-house	0	1	0.810	0	1
	(0)	(0)	(0.392)		
Existence of Patents					
EOP1	0.594	0.749	0.719	0	1
	(0.491)	(0.434)	(0.449)		
EOP2	3.364	11.53	9.983	0	98
	(7.358)	(17.46)	(16.36)		
EOP3	0.687	1.450	1.305	0	14.08
	(1.392)	(2.040)	(1.957)		
Length of Patents					
LOP1	6.764	11.17	10.34	0	20
	(7.299)	(8.143)	(8.174)		
LOP2 (x10)	3.152	10.92	9.444	0	102.1
	(7.347)	(17.09)	(16.01)		
LOP3 (x10)	0.724	1.476	1.333	0	17.66
	(1.620)	(2.286)	(2.196)		

Table 4: Distribution of In-house Development Across Broad Therapeutic Areas 1989-2004

	Mean	Std.Dev	Count
alimentary/metabolic	0.838	0.369	3623
blood and clotting	0.875	0.331	2593
cardiovascular	0.889	0.315	6262
dermatological	0.761	0.426	1261
formulations	0.703	0.457	7914
genitourinary	0.814	0.390	1320
hormonal products	0.804	0.397	632
immunological	0.774	0.419	1772
anti-infective	0.853	0.354	9552
anticancer	0.803	0.397	7849
musculoskeletal	0.838	0.368	4021
neurological	0.888	0.315	9464
antiparasitic	0.966	0.182	264
respiratory	0.869	0.338	2594
sensory	0.844	0.363	410
biotechnology	0.642	0.479	13138
miscellaneous	0.656	0.475	948
Observations	73617		

Table 5: Descriptive Statistics on the Control Variables 1989-2004

Level of Observation	Variable	Obs.	Mean	Std.Dev	Min.	Max.
Compound-year	Pre-clinical	64505	0.842	0.365	0	1
	Phase I	64505	0.0623	0.242	0	1
	Phase II	64505	0.0957	0.294	0	1
	Phase III	64505	0	0	0	0
	Launched	64505	0	0	0	0
Firm-class-year	Experience(x10)	16993	2.665	6.625	0.100	108.3
	PDM	16993	19.65	18.17	0	86
Firm-year	Scope	5695	0.626	0.324	0.100	1
Class-year	TDM	272	11.57	14.28	0	86

Table 6: Logit Models of In-house Development: Existence of Patents

	(1) In-house	(2) In-house	(3) In-house	(4) In-house	(5) In-house	(6) In-house
EOP1	0.641***	0.350***				
	(0.0551)	(0.0600)				
EOP2			0.0603***	0.0521***		
			(0.00519)	(0.00585)		
EOP3					0.243***	0.101***
					(0.0239)	(0.0302)
Phase I	-0.774***	-0.748***	-0.719***	-0.718***	-0.751***	-0.739***
	(0.0718)	(0.0714)	(0.0704)	(0.0704)	(0.0714)	(0.0711)
Phase II	-1.123***	-1.125***	-1.087***	-1.089***	-1.127***	-1.121***
	(0.0677)	(0.0680)	(0.0683)	(0.0685)	(0.0682)	(0.0681)
Experience		0.00227***		0.000333		0.00193***
		(0.000295)		(0.000287)		(0.000381)
Scope		-0.528***		-0.445***		-0.553***
		(0.0992)		(0.0972)		(0.0982)
Constant	1.534***	1.977***	1.540***	1.751***	2.002***	2.222***
	(0.178)	(0.182)	(0.178)	(0.182)	(0.172)	(0.174)
Observations	64505	64505	64505	64505	64505	64505

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators. EOP1 is an indicator for at least one other patented compound in the same therapeutic class and same firm as the compound of interest. EOP2 is the number of other patented compounds in the same therapeutic class and same firm. EOP3 is the expected number of other patented compounds in the same therapeutic class and same firm that would reach the market.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 7: Logit Models of In-house Development: Length of Patents

	(1) In-house	(2) In-house	(3) In-house	(4) In-house	(5) In-house	(6) In-house
LOP1	0.0575***	0.0410***				
	(0.00354)	(0.00445)				
LOP2			0.00581***	0.00514***		
			(0.000502)	(0.000608)		
LOP3					0.0183***	0.00245
					(0.00200)	(0.00243)
Phase I	-0.765***	-0.752***	-0.721***	-0.721***	-0.753***	-0.735***
	(0.0720)	(0.0717)	(0.0706)	(0.0706)	(0.0715)	(0.0711)
Phase II	-1.142***	-1.138***	-1.097***	-1.097***	-1.124***	-1.116***
	(0.0685)	(0.0684)	(0.0683)	(0.0684)	(0.0680)	(0.0679)
Experience		0.00131***		0.000152		0.00261***
		(0.000284)		(0.000315)		(0.000401)
Scope		-0.322***		-0.439***		-0.581***
		(0.103)		(0.0979)		(0.0985)
Constant	1.435***	1.729***	1.649***	1.833***	2.027***	2.251***
	(0.178)	(0.186)	(0.176)	(0.181)	(0.172)	(0.174)
Observations	64505	64505	64505	64505	64505	64505

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators. EOP1 is an indicator for at least one other patented compound in the same therapeutic class and same firm as the compound of interest. EOP2 is the number of other patented compounds in the same therapeutic class and same firm. EOP3 is the expected number of other patented compounds in the same therapeutic class and same firm that would reach the market.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 8: Logit Models of In-house Development: Existence of Patents with Interaction

	(1) In-house	(2) In-house	(3) In-house
EOP1	0.617***		
	(0.0909)		
EOP1 × PDM	-0.0108***		
	(0.00259)		
EOP2		0.0596***	
		(0.00928)	
EOP2 × PDM		-0.000347	
		(0.000364)	
EOP3			0.378***
			(0.0622)
EOP3 × PDM			-0.00178**
			(0.000897)
PDM	0.0243	0.0657**	0.316***
	(0.0297)	(0.0311)	(0.0671)
TDM	-0.0115	-0.0627**	-0.309***
	(0.0296)	(0.0312)	(0.0676)
Experience	0.00241***	0.000696**	0.00149***
	(0.000334)	(0.000318)	(0.000359)
Scope	-0.495***	-0.469***	-0.584***
	(0.100)	(0.0975)	(0.0974)
Phase I	-0.752***	-0.716***	-0.741***
	(0.0716)	(0.0703)	(0.0709)
Phase II	-1.125***	-1.083***	-1.119***
	(0.0681)	(0.0685)	(0.0683)
Constant	1.841***	1.743***	2.252***
	(0.194)	(0.192)	(0.178)
Observations	64505	64505	64505

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 9: Logit Models of In-house Development: Length of Patents with Interaction

	(1) In-house	(2) In-house	(3) In-house
LOP1	0.0675***		
	(0.00628)		
LOP1 × PDM	-0.00117***		
	(0.000185)		
LOP2		0.00681***	
		(0.000942)	
LOP2 × PDM		-0.0000885***	
		(0.0000317)	
LOP3			0.0153***
			(0.00535)
LOP3 × PDM			-0.000267***
			(0.0000713)
PDM	0.0436	0.0467	0.0634
	(0.0282)	(0.0316)	(0.0656)
TDM	-0.0316	-0.0417	-0.0539
	(0.0283)	(0.0317)	(0.0658)
Experience	0.00168***	0.000628*	0.00235***
	(0.000318)	(0.000345)	(0.000406)
Scope	-0.325***	-0.467***	-0.590***
	(0.102)	(0.0977)	(0.0981)
Phase I	-0.749***	-0.720***	-0.742***
	(0.0718)	(0.0705)	(0.0711)
Phase II	-1.132***	-1.091***	-1.119***
	(0.0687)	(0.0685)	(0.0681)
Constant	1.542***	1.799***	2.318***
	(0.196)	(0.188)	(0.178)
Observations	64505	64505	64505

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 10: Logit Models of In-house Development: Market Share and Existence of Patents

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
Panel A				
Current MSP	-0.905	-1.431	-0.627	-2.982
	(2.320)	(2.168)	(2.241)	(1.938)
EOP1		0.565***		
		(0.102)		
EOP2			0.0508***	
			(0.00799)	
EOP3				0.347***
				(0.0610)
Constant	2.707***	2.255***	2.148***	2.461***
	(0.157)	(0.177)	(0.172)	(0.162)
Observations	35137	35137	35137	35137
Panel B				
Future MSP	22.36**	19.75**	20.79**	21.39**
	(9.013)	(8.455)	(8.648)	(8.894)
EOP1		0.670***		
		(0.187)		
EOP2			0.0371***	
			(0.0133)	
EOP3				0.127
				(0.176)
Constant	3.136***	2.690***	2.902***	3.104***
	(0.378)	(0.389)	(0.382)	(0.384)
Observations	7009	7009	7009	7009

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include experience, scope, development phase indicators, as well as a full set of therapeutic category and year indicators.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 11: Logit Models of In-house Development: Market Share and Length of Patents

	(1) In-house	(2) In-house	(3) In-house
Panel A			
Current MSP	-2.195	-0.963	-2.547
	(2.016)	(2.171)	(1.984)
LOP1	0.0642***		
	(0.00719)		
LOP2		0.00488***	
		(0.00082)	
LOP3			0.0205***
			(0.00477)
Constant	1.856***	2.256***	2.607***
	(0.185)	(0.170)	(0.158)
Observations	35137	35137	35137
Panel B			
Future MSP	19.37**	22.12**	22.22**
	(8.098)	(8.900)	(9.076)
LOP1	0.060***		
	(0.0125)		
LOP2		0.00287**	
		(0.00132)	
LOP3			0.00195
			(0.0139)
Constant	2.477***	2.983***	3.133***
	(0.398)	(0.379)	(0.381)
Observations	7009	7009	7009

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include experience, scope, development phase indicators, as well as a full set of therapeutic category and year indicators.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 12: Logit Models of Alternative In-house Development: Existence of Patents

	(1) Alt In-house	(2) Alt In-house	(3) Alt In-house
EOP1	0.375***		
	(0.0630)		
EOP2		0.0589***	
		(0.00657)	
EOP3			0.108***
			(0.0342)
Experience	0.00229***	0.000115	0.00195***
	(0.000330)	(0.000313)	(0.000434)
Scope	-0.449***	-0.355***	-0.479***
	(0.104)	(0.101)	(0.102)
Phase I	-0.749***	-0.718***	-0.739***
	(0.0713)	(0.0703)	(0.0709)
Constant	1.908***	1.656***	2.178***
	(0.191)	(0.193)	(0.182)
Observations	58331	58331	58331

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 13: Logit Models of Alternative In-house Development: Length of Patents

	(1) Alt In-house	(2) Alt In-house	(3) Alt In-house
LOP1	0.0444***		
	(0.00477)		
LOP2		0.00584***	
		(0.00069)	
LOP3			0.00191
			(0.00269)
Experience	0.00123***	-0.0000795	0.00277***
	(0.000315)	(0.000347)	(0.000457)
Scope	-0.226***	-0.348***	-0.511***
	(0.108)	(0.102)	(0.103)
Phase I	-0.755***	-0.720***	-0.734***
	(0.0716)	(0.0705)	(0.0709)
Constant	1.631***	1.746***	2.210***
	(0.195)	(0.191)	(0.182)
Observations	58331	58331	58331

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 14: Patent Profile Variables Defined on Finer Therapeutic Classifications

	(1) In-house	(2) In-house	(3) In-house	(4) In-house	(5) In-house	(6) In-house
EOP1	0.567*** (0.0641)					
EOP2		0.185*** (0.0226)				
EOP3			0.283*** (0.0806)			
LOP1				0.0562*** (0.00593)		
LOP2					0.00201*** (0.00255)	
LOP3						0.0152** (0.00666)
Experience	0.00387*** (0.000894)	0.00117 (0.000973)	0.00525*** (0.00108)	0.00220** (0.000949)	0.00121 (0.00105)	0.00627*** (0.000106)
Scope	-0.827*** (0.114)	-0.799*** (0.114)	-0.895*** (0.115)	-0.738*** (0.115)	-0.760*** (0.115)	-0.887*** (0.115)
Phase I	-0.818*** (0.0720)	-0.789*** (0.0714)	-0.804*** (0.0718)	-0.807*** (0.0720)	-0.789*** (0.0714)	-0.803*** (0.0717)
Phase II	-1.204*** (0.0703)	-1.174*** (0.0705)	-1.197*** (0.0699)	-1.204*** (0.0706)	-1.183*** (0.0705)	-1.195*** (0.0699)
Constant	1.371*** (0.244)	1.409*** (0.242)	1.730*** (0.238)	1.380*** (0.242)	1.489*** (0.240)	1.726*** (0.238)
Observations	63317	63317	63317	63317	63317	63317

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Appendix A Proofs

Proof of Lemma 1. In any period t in which only the existing product is available, i.e., $t < 3$, each consumer maximizes

$$U(Q) = \int_0^Q (V^+ - vq)dq - (P + ds)Q \tag{A.1}$$

with respect to the quantity Q.[39] If more than one product is available, the consumer maximizes Equation (A.1) once he or she has determined which of the two products she buys. This optimal quantity is

$$\boldsymbol{Q}(P, s) = \frac{V^+ - P - ds}{v}, \tag{A.2}$$

leading to an equilibrium utility of

$$U(\boldsymbol{Q}) = \frac{(V^+ - P - ds)^2}{2v}, \tag{A.3}$$

where boldface characters indicate equilibrium values. Observe from Equation (A.3) that a consumer chooses product i over product j whenever

$$P_i + ds_i < P_j + ds_j, \tag{A.4}$$

where s_i refers to the distance of the consumer from product i in product space.

Note that the originator will always choose the monopoly price for the existing product in periods t, $t < 3$, and that the originator can only earn profits from the new product in periods t after which the existing product's patent has expired, $t > t_E$.

As such, the most interesting problem concerns periods t, $2 < t \leq t_E$, in which both products will be sold at a profit. For a given location l for the new product and prices P_O (for the originator's existing product) and P_I (for the new product or innovation), profits can be defined by means of the indifferent consumers on both sides of either product. Let

$$\theta_{OI}(P_O, P_I) = \frac{P_I - P_O}{2d} + \frac{l}{2} \tag{A.5}$$

denote the indifferent consumer located between 0 and l and

$$\theta_{IO}(P_O, P_I) = \frac{P_O - P_I}{2d} + \frac{1}{2} + \frac{l}{2} \tag{A.6}$$

the indifferent consumer located between l and 1. Then, profits accrued by product O are

[39] Recall that we assume V^+ to be large enough such that it is always optimal to serve the entire market.

given by

$$\Pi_O(P_O, P_I) = \left[\int\limits_0^{\theta_{OI}(P_O,P_I)} \boldsymbol{Q}(P_O, y)\, dy + \int\limits_0^{1-\theta_{IO}(P_O,P_I)} \boldsymbol{Q}(P_O, y)\, dy \right] (P_O - c)^{40} \tag{A.7}$$

while product I's profits are

$$\Pi_I(P_O, P_I) = \left[\int\limits_0^{l-\theta_{OI}(P_O,P_I)} \boldsymbol{Q}(P_I, y)\, dy + \int\limits_0^{\theta_{IO}(P_O,P_I)-l} \boldsymbol{Q}(P_I, y)\, dy \right] (P_I - c). \tag{A.8}$$

At the very beginning of each period t, $2 < t \leq T$, both players can make suggestions about who is assigned production, sales and pricing rights, and the (potentially negative) transfer from the originator to the licensee. These proposals may condition on the quantity of the new product sold each period but not on the exact price or location in product space due to inherent non-verifiability. The sequence of proposals in any given period is immaterial. A subgame-perfect Nash equilibrium featuring a renegotiation-proof contract demands that there is an equilibrium in behavioral strategies that coincides with the equilibrium strategies chosen at $t = 0$. Moreover, it requires that there is no equilibrium in behavioral strategies that constitutes an alternative contract at any $t > 0$.

The first result that can be established is that production of the new product in any Nash equilibrium of the game is always assigned to the developer in all periods since

$$c^+ > c. \tag{A.9}$$

Suppose this is not true in period t. Then, either firm could suggest at the beginning of period t to reassign production—immaterial to incentives—and split the additional profits in any interior way while adhering otherwise to the original contract. This proves the production part of both i) and ii). Moreover, we can establish that, for every period t, $t > 2$, sales and pricing rights are allocated in a renegotiation-proof subgame-perfect Nash equilibrium in such a way as to maximize total surplus in a given period. For if not, a Pareto-improving renegotiation is possible.

This immediately implies that sales and pricing rights for the new product can only ever be allocated to the licensee if the licensee internalizes variable cost and revenue of all products held by the originator in a given period. It follows that sales and pricing rights at t, $2 < t \leq t_E$, necessarily reside with the originator. This establishes the sales and pricing part of i). That is to say, the originator solves

$$\max_{P_O,P_I} \Pi_{OI}(p_O, p_I) = \Pi_O(p_O, p_I) + \Pi_I(p_O, p_I) \tag{A.10}$$

in periods 3 to t_E.

Clearly, the overall profit in any given period depends on the location of the new product

[40] Through this section we set $c \equiv c_1 = c_2$ significantly simplifying algebra without affecting any of the qualitative results presented in this paper.

l. Π_{OI} can be shown to have three local maxima in the (P_O, P_I) space. The global maximum, however, is the same for each $l \in [0, 1)$. Equilibrium prices in period t, $2 < t \leq t_E$, are given by

$$\boldsymbol{P_O} = \boldsymbol{P_I} = \frac{4(V^+ - c) - d + 2dl(1-l)}{8} \tag{A.11}$$

for any realized location of the new product l. As such,

$$\Pi_{OI}(\boldsymbol{P_O}, \boldsymbol{P_I}) = \frac{\big(4(V^+ - c) - d + 2dl(1-l)\big)^2}{64v}, \tag{A.12}$$

and

$$\frac{\partial \Pi_{OI}(\boldsymbol{P_O}, \boldsymbol{P_I})}{\partial l} = 0 \tag{A.13}$$

at $l = \frac{1}{2}$ only with

$$\frac{\partial^2 \Pi_{OI}(\boldsymbol{P_O}, \boldsymbol{P_I})}{\partial^2 l} < 0 \tag{A.14}$$

if V^+ is large enough. It follows that the originator strictly prefers $l = \frac{1}{2}$.

If the licensee, however, never internalizes variable costs and revenue, it does not care about the location of the new product. Therefore it chooses $k = 0$. Can the originator do better? Indeed, it can.

By allocating sales and pricing rights of the new product in periods t, $t > t_E$, the originator forces the licensee to internalize variable cost and revenue and thus maximize the profits of the new product. This induces the licensee to care about the location of the new product. Once the patent of the originator's existing product has expired, product O is sold at price c. For large enough V^+ the licensee chooses to price the new product at

$$\frac{8V^+ + 10c + 2d + \sqrt{64V^{+2} - 128cV^+ + 64c^2 + 40cd - 40dV^+ + 22d^2 - 36d^2l(1-l)}}{18} \tag{A.15}$$

and its profits are uniquely maximized at $l = \frac{1}{2}$. As a consequence, the licensee chooses $k > 0$ in period 1 as $p'(0) = \infty$.

Finally, there is a competitive pool of licensees. As such, the licensee necessarily expects to make zero economic profit ex ante. This means that for each period t, $2 < t \leq t_E$, the originator guarantees to pay the licensee c times the quantity of the new product sold. Moreover, the the licensee agrees to pay to the originator the expected profit from the new product in periods $t_E + 1$ to T minus the incurred fixed cost F_L. This establishes the remainder of i) and ii) as well as iii). ■

Proof of Lemma 2. We have established in the proof of Lemma 1 above that the originator always prefers to locate the new product at $l = \frac{1}{2}$ in periods t, $2 < t \leq t_E$. The same is true for the originator, and equivalently, for the licensee in periods $t > t_E$. i) follows.

ii) follows directly since the originator and a potential licensee face the same optimization problem in periods $t > t_E$. If $t_E < 3$, the existing and new product are never sold under a patent in the same period and only periods t, $t > t_E$, affect the incentives of the developer to invest in location precision.

The originator and licensee have the same incentive to invest in location precision for all periods t, $t > t_E$. However, from the proof of Lemma 1 it is clear that the originator faces an additional incentive to invest in location precision to increase its profits through period t_E. Moreover, if $K(O,T) = 0$, the licensee never controls pricing and sales and thus never internalizes variable costs or revenue. As a result, the licensee has no incentive to invest. This completes the proof of iii).

The later the patent of the existing product expires, the longer the originator as developer invests in maximizing expected distance between the two products for the sake of both products' profits. By the discussion above, it follows that this investment increases in the number of periods with a valid patent. The last step follows from ii) above. This argument proves iv).

Finally, the reverse is true if the licensee is developing the product. The licensee's outcome only depends on profits in periods $t > t_E$. Therefore, the licensee invests more in maximizing expected distance between the products, the more periods it benefits from profits. This establishes v). ■

Proof of Proposition 1. We will prove three conditions, which when combined establish the claim. First, we will show that, for every vector of admissible parameters, there is an equilibrium of the subgame that is initiated when the originator chooses to outsource development of the new product. Second, the originator's expected equilibrium profit in this subgame is unique. And, if an equilibrium of the one-player subgame in which the originator develops the new product internally exists, it is unique as well. Finally third, there is a unique subgame perfect equilibrium of the entire game as it pertains to expected outcomes for both parties and this equilibrium is a cutoff equilibrium of the form described in the claim of the proposition.

First, by assumption we focus on parameter values under which all consumers buy either of the two products in every period and it is profitable to have a licensee develop the new product even if the licensee were to choose an investment level $k = 0$. We know that whenever development is outsourced, in this subgame the licensee controls production while sales and pricing rights are in the hands of the originator for t, $2 < t \leq t_E$, and under the control of the licensee for periods t, $t > \max\{2, t_E\}$. To ensure the existence of an equilibrium in this subgame, we have to establish that there is no vector of parameters such that allocating sales and pricing in any period $t \leq t_E$ to the licensee results in higher total surplus by motivating the licensee to choose a more efficient k. While Lemma 1 i) shows that such a contract cannot constitute an equilibrium, we have not ruled out that, for some parameters, it may constitute a profitable deviation from the contract. Assume that sales and pricing are allocated to the licensee in some period $t \leq t_E$, and that the potential gain from a larger k outweighs the loss from price competition between the originator with the patented product and the licensee with the new product in t. This logic, however, is flawed. Once ε has been realized, there is always a follow-up contract that would make the licensee better off giving up sales and pricing rights in t with $2 < t \leq t_E$. As such, the licensee would not choose a socially better k than in the first place. It follows that this subgame always has an equilibrium.

Second, i) and ii) of Lemma 1 together with the first part of iii) of Lemma 1 establish the uniqueness of this subgame equilibrium in terms of profits, since all rights are unambiguously assigned every single period and the expected profit of the licensee equals 0. While the timing of fixed payments is innocuous as there must always be one party objecting to a contract change reducing its profits, this pins down the expected subgame equilibrium profits of the originator uniquely. Now consider the one-player subgame initiated by the originator choosing to develop the new product internally. In this scenario, the originator retains all rights for all periods and thus chooses the socially optimal $k = k^*$. As a consequence, this subgame clearly has a unique equilibrium.

Third, let the expected profit of the originator from outsourcing equal $\Pi(O)$ while the expected profit from internally developing the new product is denoted as $\Pi(I, F)$. It follows from the discussion above that for any given set of parameters, if the subgame initiated by the originator choosing to develop the new product internally has an equilibrium, $\Pi(I, F)$ equals a positive constant minus F. The assumption about the feasibility of positive profits when outsourcing coupled with the fact that the originator chooses the socially efficient $k = k^*$ implies that

$$\Pi(I, F_L) > \Pi(O) > 0. \tag{A.16}$$

Thus, by the continuity of $\Pi(I, F)$ in F, there necessarily is an $F^* > F_L$ such that

$$\Pi(I, F^*) = \Pi(O) > 0. \tag{A.17}$$

It follows that the subgame perfect equilibrium of the overall game—that is unique up to timing of transfers if development is outsourced as argued above—has the originator choose to develop the new product internally if $F \leq F^*$ and to outsource if $F > F^*$. Defining $\Delta^* = F^* - F_L$ establishes the first part of the claim. Moreover, uniqueness follows trivially.

Finally, it remains to be shown that a) $\Delta^* = 0$ for t_E, $t_E \leq 2$, and b) Δ^* is strictly increasing for t_E, $t_E \geq 2$. a) follows from Lemma 2 ii) and the fact that the expected profit of the licensee equals 0. Now consider Lemma 2, points ii), iv) and v). Together these statements imply that the k chosen for t_E, $t_E \geq 2$, by the licensee when developing the product always falls short of k^*, the optimal k as chosen by the originator when developing the new product in-house. What is more, they collectively imply that

$$\frac{\partial \left(K(I, t_E) - K(O, t_E) \right)}{\partial t_E} > 0, \tag{A.18}$$

i.e., strictly increases in t_E, $t_E \geq 2$. As a consequence, the nominal welfare loss from delegating the development to the licensee strictly increases in t_E for fixed T. This establishes b). ■

Appendix B Extensions

In this section we provide two extensions of our basic model introduced in Section 3. The first extension addresses the presence of competition and the market share of existing patents as a predictor of the outsourcing decision. The second extension concerns a single firm that faces multiple outsourcing decisions.

B.1 Competition

We formally present a 2-period model without discounting that simplifies along several dimensions, and argue why the underlying logic applies under more general circumstances. Each consumer is interested in buying a single product in each period only. Moreover, there is a now a competitor we call the rival who offers a competing product in the same product category starting in both periods.

More specifically, consider the product space to be an equilateral triangle with circumference 1, which, naturally, is homeomorphic to a circle but lends itself better for describing the outsourcing decision of an innovating firm in the presence of a competitor. Label its corners clockwise starting at the top by A, B, and C, i.e., in terms of location $A = 0$, $B = \frac{1}{3}$, and $C = \frac{2}{3}$. Consumers with a total mass of 1 are uniformly distributed along the perimeter of the triangle.

Assume the originator O to have a patented product located at A and a rival R having a patented product located at B. In period 1, only products O and R vie for the consumers located around the triangle by simultaneously setting prices.

At the end of period 1, the originator innovates and introduces a second product. We denote this product by the subscript I. We further assume that the developer can only position its product in product space between A and C due to technological feasibility or the patent that the rival holds. The product's final location is determined by the product developer's location choice as well as a random process just as in Section 3 that depends on the developer's investment in location precision k. In fact, just as before, the new product's location is $l = l_M + \varepsilon$, where ε is drawn from either of two uniform distributions $U[-\alpha, \alpha]$ and $U[-\beta, \beta]$ with $\alpha < \beta \leq \frac{1}{12}$. If the developer chooses to locate the product nearby A or B, ε is drawn from the resulting conditional distribution truncated at the respective end of the line connecting A and B. In addition, we simplify by assuming that all three products can be produced at 0 marginal cost. If, however, a firm that did not develop product I produces it, it exhibits a positive marginal cost.

Just as in Section 3 the originator chooses whether to outsource the development of product I to a firm from a competitive pool of licensees or to develop and produce in-house. This choice depends on the realization of the originator's stochastic fixed cost of development F in relation to the known licensees' fixed cost of development F_L. If the originator decides to outsource, a renegotiation-proof contract is signed in period 1 before the development of the product.

We assume that the new product is introduced after period 1, and offered alongside products O and R. The following intuition carries over from Sections 3. Naturally, if the originator decides to develop product I there is no outsourcing, and the originator obtains production, sales, and pricing rights of product I. If, however, the originator decides to outsource, the licensee is assigned production of product I, but never pricing and sales rights since product O and I are both offered in period 2 and the licensee would not internalize the effects of its pricing choice on product O.

A consumer's utility from buying product I is determined by the product's utility V_i^+ (where we assume $V_I^+ = V_O^+$, i.e., technological innovation takes place in product space, not affecting the product's base utility), its price P_i and the consumer's distance from the product's location in product space. That is, for example a consumer located at θ on the line between A and B receives a utility when buying from the originator of

$$U_O(\theta) = V_O^+ - P_O - \delta(\theta) \text{ and } U_R(\theta) = V_R^+ - P_R - \delta(\frac{1}{3} - \theta) \tag{B.1}$$

when buying from its rival, where $\delta(\cdot)$ denotes the distance cost function. Throughout this subsection we assume $\delta(\cdot)$ to be the identity function and V_O^+ and V_R^+ to be sufficiently large such that every consumer buys some product in equilibrium.

Below we show that in this environment, the originator is less likely to outsource development and production of the new product the larger its initial market share in the first period, i.e., its predicted market share in period two in the absence of innovation.

Proposition 2 *In the unique equilibrium outcome of the* 2*-period innovation game with competition, the originator is more likely to choose in-house development the higher its market share before the introduction of the new product.*

Proof of Proposition 2. It is straightforward to see that competition in period 1 in this model is equivalent to competition of two competitors with fixed positions on a Salop circle. In equilibrium, the originator chooses a price of

$$P_O^1 = \frac{V_O^+ - V_R^+}{3} + \frac{1}{2} \tag{B.2}$$

to obtain market share

$$MS_O^1 = \frac{V_O^+ - V_R^+}{3} + \frac{1}{2}, \tag{B.3}$$

such that the originator's equilibrium market share strictly increases in V_O^+ and strictly decreases in V_R^+.[41]

Let us now assume that both the originator and a licensee would locate product I at $l_M = C$, an assumption that we will justify below. Let θ_{OR} denote the consumer who is located between A and B and indifferent between buying from O and R, θ_{RI} the consumer located between B and $C + \epsilon$, where ϵ refers to the realization of ε, and indifferent between buying from R and I, and θ_{IO} the consumer located between $C + \epsilon$ and A and indifferent between buying products I or O.

In period 2, the originator maximizes

$$\begin{aligned}\Pi_{OI}(P_O^2, P_I^2, P_R^2) = &\Big[\theta_{OR}(P_O^2, P_R^2) + \big(1 - \theta_{IO}(P_I^2, P_O^2)\big)\Big] P_O^2 \\ &+ \big(\theta_{IO}(P_I^2, P_O^2) - \theta_{RI}(P_R^2, P_I^2)\big) P_I^2\end{aligned} \tag{B.4}$$

by choosing P_O^2 and P_I^2, with the superscript indicating the period.

In equilibrium, the originator chooses to price its products at

$$\boldsymbol{P_O^2} = \frac{12V_O^+ - 12V_R^+ + 20 - 9\epsilon}{36}, \tag{B.5}$$

[41] We assume throughout this subsection that $|V_O^+ - V_R^+|$ is sufficiently small such that all two (three) products exhibit positive demand in period 2 (3). We refer to such an equilibrium as an interior solution.

and

$$\boldsymbol{P_I^2} = \frac{12V_O^+ - 12V_R^+ + 20 - 3\epsilon}{36}, \tag{B.6}$$

respectively, while its rival prices at

$$\boldsymbol{P_R^2} = \frac{6V_R^+ - 6V_O^+ + 8 + 3\epsilon}{18}, \tag{B.7}$$

with bold variables indicating equilibrium values.

Substituting these equilibrium prices into Equation (B.4) and derivation with respect to location precision results in

$$\frac{\partial \Pi_{OI}(\boldsymbol{P_O^2}, \boldsymbol{P_I^2}, \boldsymbol{P_R^2})}{\partial \epsilon} = \frac{144V_R^+ - 144V_O^+ + 126\epsilon - 240}{1296}, \tag{B.8}$$

an expression that is always negative for an interior solution. This establishes that the originator prefers product I to be located at $l = C$, and, in fact, always prefers the product to be closer to C.

In addition, we can now see that

$$\frac{\partial^2 \Pi_{OI}(\boldsymbol{P_O^2}, \boldsymbol{P_I^2}, \boldsymbol{P_R^2})}{\partial \epsilon \partial (V_O^+ - V_R^+)} = -\frac{1}{9}. \tag{B.9}$$

This implies that location precision is more important for the originator's profits the larger its period 1 market share.

Since a licensee as developer does not internalize revenue or variable cost since it is paid a fixed amount, it does not invest at all in location precision, i.e., $k = 0$. It chooses, however, $l_M = C = \frac{2}{3}$ in equilibrium, as, otherwise, the originator contracts with another licensee. The originator as developer on the other hand chooses the efficient investment in location precision k^*. Note that $k^* > 0$ due to $p'(0) = \infty$ and Equation (B.8). Moreover, by the argument above based on Equation (B.9), it increases in the originator's period 1 market share.

Just as in Section 3, the payments from the originator to a developing licensee are such that the licensee earns zero economic profits, i.e. they amount to F_L in this case. Denote the originator's expected equilibrium profits in period 2 as a function of the investment level $\mathbb{E}\boldsymbol{\Pi_{OI}^2}(k)$, and note that

$$\frac{\partial \mathbb{E}\boldsymbol{\Pi_{OI}^2}(k)}{\partial k} > 0 \tag{B.10}$$

for all $k \in [0, k^*)$.

It follows that the originator chooses in-house development if

$$\mathbb{E}\boldsymbol{\Pi_{OI}^2}(k^*) - F > \mathbb{E}\boldsymbol{\Pi_{OI}^2}(0) - F_L. \tag{B.11}$$

This gives rise to a

$$\Delta^* = \mathbb{E}\boldsymbol{\Pi_{OI}^2}(k^*) - \mathbb{E}\boldsymbol{\Pi_{OI}^2}(0), \tag{B.12}$$

such that the originator chooses to outsource development if and only if

$$\Delta = F - F_L \tag{B.13}$$

exceeds Δ^*. It follows directly from Equation (B.9) that $\mathbb{E}\boldsymbol{\Pi}^{\mathbf{2}}_{\boldsymbol{OI}}(k^*)$ increases more in

$$V_O^+ - V_R^+, \tag{B.14}$$

and therefore in MS_O^1, than does $\mathbb{E}\boldsymbol{\Pi}^{\mathbf{2}}_{\boldsymbol{OI}}(k^*)$, and thus Δ^* increases in MS_O^1. This proves the claim. ■

The two-period model introduced in this subsection establishes that if both the existing as well as a new product hold patents at the same time, the originator is more likely to choose in-house development the higher the market share of its existing product in the first place, thus rationalizing Testable Prediction 3. Clearly this insight generalizes to the multiple periods setting in Section 3 as periods in which only one of the products holds a patent do not affect the rationale underlying this result. It can also be generalized in terms of the distance cost function, and for consumers who buy more than one unit of a product such as in Section 3. Moreover, the model can be extended to one in which consumers populate the interior of the triangle.

B.2 Multiple R&D investments

Consider the model presented in Section 2 and assume that the originator faces two product development tasks, tasks 1 and 2, when developing a new product, each of which may be outsourced to companies from a pool of competitive licensees, or undertaken in-house. The originator now faces two random draws from probability density functions $f^1(\cdot)$ and $f^2(\cdot)$, the realizations of which determine its fixed cost of taking on the respective task in-house. Licensees on the other hand incur known fixed costs of F_L^1, and F_L^2 respectively. Let Δ^i refer to the originator's fixed cost disadvantage related to task i. Likewise we assume that a firm that has undertaken one of the development tasks has a competitive advantage in production, i.e., its marginal production cost c satisfies $c < c^+$, with c^+ being the cost of a company that has not fulfilled either of the development tasks.

The location of the new product in product space is determined by the location choice and investment in location precision of both firms undertaking a development task. Specifically assume that both developing companies face the familiar investment decision in location precision. The more they invest, the likelier it is that the random error regarding the location of the product is drawn from a favorable distribution, i.e., $U[-\alpha, \alpha]$ instead of $U[-\beta, \beta]$. We continue to assume $\alpha < \beta \leq \frac{1}{2} - (c - c_1)$. The final location of the new product in product space is then determined by the convex combination

$$l = \phi_1 \left(l_M^1 + \varepsilon^1\right) + \phi_2 \left(l_M^2 + \varepsilon^2\right), \tag{B.15}$$

where $\phi_i > 0$, $\phi_1 + \phi_2 = 1$, denotes the contribution of task i to the final location of the product. Furthermore, assume that both firms make their location and investment decisions at the same time that the originator decides who controls production and pricing.

If, in equilibrium, the originator ends up outsourcing either none or one of the tasks, the equilibrium solution and contracts are—bar minor details—given by the solutions presented in Sections 3 and 4. As a consequence, we focus in this subsection on the equilibrium outcome in which the originator outsources both development tasks to licensees and argue under which circumstances this outcome arises. Note that the basic intuition from Section 3 about contracts carries over. As such, it has to be true that either of the licensees (we refer to them as licensee 1 and licensee 2 according to their task), but not the originator, undertakes production of the new product. Furthermore, the originator will not delegate pricing rights to licensees before the patent of its existing product expires. Which licensee will be allocated the right to price the new product and collect revenue?

Proposition 3 *In the unique equilibrium of the innovation game in which the originator chooses to outsource multiple tasks, the originator assigns production and pricing rights to each licensee with positive probability. Moreover, the originator is less likely to outsource a development task i) if it has an existing patented product in the same product class, and ii) the longer this patent is expected to be valid after the new product reaches the market.*

Proof of Proposition 3. First consider the case in which the originator outsources both tasks but assigns pricing w.l.o.g. with certainty to licensee 1. In this scenario, licensee 2 will not invest at all in location precision. In this scenario, the originator understands that as

$$p'(0) = \infty,\ p'(k) > 0\ \forall k > 0,\ p(\infty) < 1 \text{ and } \phi_i > 0 \text{ for } i \in \{1, 2\}, \tag{B.16}$$

assigning production and pricing of the new product with an infinitesimal probability to licensee 2 pushes l in expectation towards $\frac{1}{2}$, the originator's strictly preferred location of the new product. This follows since licensee 2 prefers the location $\frac{1}{2}$ if there is any chance that it collects the revenue of the new product for at least some periods while paying a fixed amount to the originator. Thus, and due to $p'(0) = \infty$, licensee 2 invests a positive amount in location precision. As $p(\cdot)$ is continuously differentiable and $p'(0) = \infty$, there is $\zeta > 0$ small enough such that

$$\phi_2 * p'(\zeta) > \phi_1 p'(K_1) \tag{B.17}$$

for any positive investment level K_1 of licensee 1 and any $\phi_2 > 0$. This establishes that the originator assigns production and pricing to both licensees with positive likelihood.

The originator never assigns task i to a licensee if

$$F^i < F^i_L \Leftrightarrow \Delta^i < 0. \tag{B.18}$$

However, even if $\Delta^i > 0$, the originator does not necessarily outsource task i. This follows from the argument in Section 3 laying out that even if licensee i chooses the optimal location of $\frac{1}{2}$ it does in general not choose the investment level preferred by the originator as it does not internalize the cannibalization of the existing product.

Naturally, if the patents of the existing and the new product do not overlap, the originator's and the licensee's objectives align perfectly. This establishes i). Moreover, just as in the base model outlined in Section 3, the originator is more likely to assign development task i to a licensee when $\Delta^i > 0$ and there is smaller number of periods in which the licensee does not internalize cannibalization, i.e., if the patent length of the existing product is shorter, establishing ii). ■

References

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and Innovation: An Inverted U-Relationship. *Quarterly Journal of Economics 120*, 701–728.

Aghion, P. and J. Tirole (1994). The Management of Innovation. *Quarterly Journal of Economics 109*, 1185–1209.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics 84*, 488–500.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In Universities-National Bureau Committee for Economic Research, Committee on Economic Growth of the Social Science Research Council (Ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626. NJ: Princeton University Press.

Azoulay, P. (2004). Capturing Knowledge Within and Across Firm Boundaries: Evidence from Clinical Development. *American Economic Review 94*, 1591–1612.

Balakrishnan, S. and B. Wernerfelt (1986). Technical Change, Competition and Vertical Integration. *Strategic Management Journal 7*, 347–359.

Barrera, C. and M. Waldman (2019). Vertical Integration, Secrecy, and High Tech Industries. *mimeo*.

Budish, E., B. N. Roin, and H. Williams (2015). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *American Economic Review 105*, 2044–2085.

Casas-Arce, P., T. Kittsteiner, and F. A. Martinez-Jerez (2019). Contracting with Opportunistic Partners: Theory and Application to Technology Development and Innovation. *Management Science 65*, 842–858.

Choi, J. P. (1994). Network Externality, Compatibility Choice, and Planned Obsolescence. *The Journal of Industrial Economics 42*, 167–182.

Cunningham, C., F. Ederer, and S. Ma (2021). Killer Acquisitions. *Journal of Political Economy 129*, 649–702.

Danzon, P. M., S. Nicholson, and N. S. Pereira (2005). Productivity in Pharmaceutical–Biotechnology R&D: The Role of Experience and Alliances. *Journal of Health Economics 24*, 317–339.

Demougin, D. M. (1989). A Renegotiation-Proof Mechanism for a Principal-Agent Model with Moral Hazard and Adverse Selection. *RAND Journal of Economics 20*, 256–266.

Dewatripont, M. (1988). Commitment Through Renegotiation-Proof Contracts with Third Parties. *Review of Economic Studies 55*, 377–390.

DiMasi, J. A., R. W. Hansen, and H. G. Grabowski (2003). The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics 22*, 151–185.

Galdon-Sanchez, J. E., R. Gil, and A. Bayo-Mariones (2015). Outsourcing of Peripheral Services: Evidence from Spanish Manufacturing Plant-Level Data. *European Economic Review 78*, 328–344.

Gans, J. S. and S. Stern (2000). Incumbency and R&D Incentives: Licensing the Gale of Creative Destruction. *Journal of Economics & Management Strategy 9*, 485–511.

Gibbons, R. (2005). Four Formal(izable) Theories of the Firm? *Journal of Economic Behavior & Organization 58*, 200–245.

Gil, R. and C. A. Ruzzier (2018). The Impact of Competition on "Make-or-Buy" Decisions: Evidence from the Spanish Local TV Industry. *Management Science 64*, 1121–1135.

Grossman, S. J. and O. D. Hart (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy 94*, 691–719.

Hart, O. (2017). Incomplete Contracts and Control. *American Economic Review 107*, 1731–1752.

Hart, O. and J. Moore (1988). Incomplete Contracts and Renegotiation. *Econometrica 56*, 755–785.

Hart, O. and J. Moore (1990). Property Rights and the Nature of the Firm. *Journal of Political Economy 98*, 1119–1158.

Higgins, M. J. and D. Rodriguez (2006). The Outsourcing of R&D Through Acquisitions in the Pharmaceutical Industry. *Journal of Financial Economics 80*, 351–383.

Igami, M. (2017). Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981-1998. *Journal of Political Economy 125*, 798–847.

Klein, B. R., R. G. Crawford, and A. A. Alchien (1978). Vertical Integration, Appropriable Rents, and the Competitive Contracting Process. *Journal of Law and Economics 21*, 297–326.

Krieger, J., D. Li, and D. Papanikolaou (2017). Developing New Drugs. *mimeo*.

Lafontaine, F. and M. Slade (2007). Vertical Integration and Firm Boundaries: The Evidence. *Journal of Economic Literature 45*, 629–685.

Lakdawalla, D. N. (2018). Economics of the Pharmaceutical Industry. *Journal of Economic Literature 56*, 397–449.

Lam, M. D. (2004, June 1). Why Alliances Fail. *Pharmaceutical Executive*.

Lerner, J. and U. Malmendier (2010). Contractibility and the Design of Research Agreements. *American Economic Review 100*, 214–246.

Lerner, J. and R. P. Merges (1998). The Control of Technology Alliances: An Empirical Analysis of the Biotechnology Industry. *The Journal of Industrial Economics 46*, 125–156.

Levy, D. T. (1985). The Transactions Cost Approach to Vertical Integration: An Empirical Examination. *The Review of Economics and Statistics 67*, 438–445.

Moorthy, K. S. and I. P. L. Png (1992). Market Segmentation, Cannibalization, and the Timing of Product Introductions. *Management Science 38*, 345–359.

Nahm, J. (2004). Durable-Goods Monopoly with Endogenous Innovation. *Journal of Economics & Management Strategy 13*, 303–319.

Nicholson, S., P. M. Danzon, and J. McCullough (2005). Biotech-Pharmaceutical Alliances as a Signal of Asset and Firm Quality. *Journal of Business 78*, 1433–1464.

Novak, S. and S. Stern (2009). Complementarity Among Vertical Integration Decisions: Evidence from Automobile Product Development. *Management Science 55*, 311–332.

Pan, Y. (2016). *Essays in Industrial Organization and Health Economics.* Dissertation: Cornell University.

Patel, D. (2017, July 17). The Pros and Cons of Outsourcing. *Forbes*.

Pisano, G. P. (1990). The R&D Boundaries of the Firm: An Empirical Analysis. *Administrative Science Quarterly 35*, 153–176.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *The Bell Journal of Economics 10*, 141–156.

Siebert, R. (2015). Entering New Markets in the Presence of Competition: Price Discrimination versus Cannibalization. *Journal of Economics & Management Strategy 24*, 369–389.

Tucker, I. B. and R. P. Wilder (1977). Trends in Vertical Integration in the U.S. Manufacturing Sector. *The Journal of Industrial Economics 26*, 81–94.

Waldman, M. (1993). A New Perspective on Planned Obsolescence. *Quarterly Journal of Economics 108*, 273–283.

Waldman, M. (1996). Planned Obsolescence and The R&D Decision. *RAND Journal of Economics 27*, 583–595.

Williams, H. L. (2013). Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy 121*, 1–27.

Williamson, O. E. (1979). Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics 22*, 233–261.