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.[36] 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

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

given by

$$\Pi_O(P_O, P_I) = \left[\int_0^{\theta_{OI}(P_O,P_I)} \boldsymbol{Q}(P_O, y)\, dy + \int_0^{1-\theta_{IO}(P_O,P_I)} \boldsymbol{Q}(P_O, y)\, dy \right] (P_O - c)^{37} \tag{A.7}$$

while product I's profits are

$$\Pi_I(P_O, P_I) = \left[\int_0^{l-\theta_{OI}(P_O,P_I)} \boldsymbol{Q}(P_I, y)\, dy + \int_0^{\theta_{IO}(P_O,P_I)-l} \boldsymbol{Q}(P_I, y)\, dy \right] (P_I - c). \tag{A.8}$$

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

[37] 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 at 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 be always 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 $\Pi(I, F_O)$. 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, $\Pi(I, F_O)$ equals a positive constant minus F_O. 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 $\Pi(I, F_O)$ in F_O, 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 $F_O \leq F^*$ and to outsource if $F_O > F^*$. 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 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's 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 but 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_O 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 Section 2. 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 a 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^+.[38]

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

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

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

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

[38] 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 makes 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

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

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

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

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

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

and therefore in MS_O^1, than does $\mathbb{E}\mathbf{\Pi}^{\mathbf{2}}_{\boldsymbol{OI}}(k^*)$, 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 developing 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 $\alpha < \beta \leq \frac{1}{4}$. 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 at which 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

$$F_O^i < F_L^i \Leftrightarrow \Delta^i < 0. \tag{B.18}$$

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 the 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) ■

The intuition derived from the model outlined in this subsection in which the originator faces two outsourcing decisions generalizes to any finite number of development tasks $N < \infty$.

Appendix C Tables

Table C.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 C.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 C.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 C.4: 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 C.5: 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 C.6: 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 C.7: 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 C.8: 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 C.9: 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 C.10: 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 (2.016)	-0.963 (2.171)	-2.547 (1.984)
LOP1	0.0642*** (0.00719)		
LOP2		0.00488*** (0.00082)	
LOP3			0.0205*** (0.00477)
Constant	1.856*** (0.185)	2.256*** (0.170)	2.607*** (0.158)
Observations	35137	35137	35137
Panel B			
Future MSP	19.37** (8.098)	22.12** (8.900)	22.22** (9.076)
LOP1	0.060*** (0.0125)		
LOP2		0.00287** (0.00132)	
LOP3			0.00195 (0.0139)
Constant	2.477*** (0.398)	2.983*** (0.379)	3.133*** (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 C.11: 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 C.12: 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 C.13: 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***	0.00117	0.00525***	0.00220**	0.00121	0.00627***
	(0.000894)	(0.000973)	(0.00108)	(0.000949)	(0.00105)	(0.000106)
Scope	-0.827***	-0.799***	-0.895***	-0.738***	-0.760***	-0.887***
	(0.114)	(0.114)	(0.115)	(0.115)	(0.115)	(0.115)
Phase I	-0.818***	-0.789***	-0.804***	-0.807***	-0.789***	-0.803***
	(0.0720)	(0.0714)	(0.0718)	(0.0720)	(0.0714)	(0.0717)
Phase II	-1.204***	-1.174***	-1.197***	-1.204***	-1.183***	-1.195***
	(0.0703)	(0.0705)	(0.0699)	(0.0706)	(0.0705)	(0.0699)
Constant	1.371***	1.409***	1.730***	1.380***	1.489***	1.726***
	(0.244)	(0.242)	(0.238)	(0.242)	(0.240)	(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.